Final Report Key Contents

Main Results Accomplished by the EU-Funded Project IM-CLeVeR – Intrinsically Motivated Cumulative Learning Versatile Robots

Authors:

Gianluca Baldassarre, Marco Mirolli, Elisabetta Visalberghi, Eugenio Guglielmelli, Flavio Keller, Peter Redgrave, Kevin Gurney, Jochen Triesch, Martin Mc Guinnity, Mark Lee, Juergen Schmidhuber, Andrew Barto

2013

An EU funded Integrated Project (IP), Grant Agreement number 231722, Duration of project: 1st May 2009 – 30th April 2013

Website: www.im-clever.eu

Coordinator contact: gianluca.baldassarre@istc.cnr.it

Guide to the contents

This document has the goal of presenting the main scientific and technological achievements of the project IM-CLeVeR. The document is organised as follows:
1. Project executive summary: a brief overview of the project vision, objectives and keywords.
2. Beneficiaries of the project and contacts: list of Teams (partners) of the project, Team Leaders and contacts.
3. Project context and objectives: the vision of the project and its overall objectives
4. Overview of work performed and main results achieved: a one page overview of the main results of the project
5. Overview of main results per partner: a bullet-point list of main results per partners
6. Main achievements in detail, per partner: a throughout explanation of the main results per partner (but including collaboration work), with also reference to the main publications supporting them.

1. Project executive summary

IM-CLeVeR aimed to develop a new methodology for designing robots controllers that can: (1) cumulatively learn new efficient skills through autonomous development based on intrinsic motivations, and (2) reuse such skills for accomplishing multiple, complex, and externally-assigned tasks. During skill-acquisition, the robots behave like children at play which acquire skills autonomously on the basis of "intrinsic motivations". During skill-exploitation, the robots exhibits fast learning capabilities and high versatility in solving tasks defined by external users due to the capacity of flexibly re-using, composing and re-adapting previously acquired skills.

This overall goal was pursued investigating three fundamental scientific and technological issues: (1) the mechanisms of abstraction of sensory information; (2) the mechanisms underlying intrinsic motivations, e.g. "curiosity drives" that learn to focus attention and learning capabilities on "zones of proximal development"; (3) hierarchical recursive architectures which permit cumulative learning. The study of these issues fuelled by a reverse-engineering effort aiming at reproducing with bio-mimetic models the results of empirical experiments run with monkeys, children, and human adults. The controllers proposed were validated with challenging demonstrators based on the iCub humanoid robotic platform.

The project significantly advanced the scientific and technological state of the art, both in terms of theory and implementations, in autonomous learning systems and robots. This goal was achieved on the basis of the integrated work of a highly interdisciplinary Consortium involving leading international neuroscientists, psychologists, roboticists and machine-learning researchers.

Project keywords
Developmental robotics, autonomous learning, reinforcement learning, action hierarchies, neural networks, novelty detection, developmental/cognitive psychology, brain, dopamine.

2. Beneficiaries of the project and contacts

Beneficiary Number	Beneficiary name	Benefic. acronym	Country
1 Coordinator	**Consiglio Nazionale delle Ricerche,** **Istituto di Scienze e Tecnologie della Cognizione:** 1.1 Laboratory of Computational Embodied Neuroscience *Coordinator*: *Gianluca Baldassarre (gianluca.baldassarre@istc.cnr.it)* 1.2 Unit of Cognitive Primatology Team Leader: *Elisabetta Visalberghi (elisabetta.visalberghi@istc.cnr.it)*	CNR-ISTC-LOCEN CNR-ISTC-UCP	Italy
2	**Universita' Campus Bio-Medico** 2.1 Lab. of Biomedical Robotics and Biomicrosystems Team Leader: *Eugenio Guglielmelli (e.guglielmelli@unicampus.it)* 2.2 Laboratory of Developmental Neuroscience Team Leader: *Flavio Keller (f.keller@unicampus.it)*	UCBM-LBRB UCBM-LDN	Italy
3	**University of Sheffield,** **Department of Psychology** Team Leader: *Peter Redgrave (peter.redgrave@usfd.ac.uk)*	USFD	United Kingdom
4	**Goethe University,** **Frankfurt Institute for Advanced Studies** Team Leader: *Jochen Triesh (triesch@fias.uni-frankfurt.de)*	FIAS	Germany
5	**University of Ulster,** **School of Computing and Intelligent Systems** Team Leader: *Martin T. McGinnity (TM.McGinnity@ulster.ac.uk)*	UU	United Kingdom
6	**Aberystwyth University,** **Department of Computer Science** Team Leader: *Mark Lee (mhl@aber.ac.uk)*	AU	United Kingdom
7	**Istituto Dalle Molle per l'Intelligenza Artificiale,** **Scuola Univ. Profes. della Svizzera Italiana** Team Leader: *Juergen Schmidhuber (juergen@idsia.ch)*	IDSIA-SUPSI	Switzerland

Legal/administrative details of the project

Project Officer: Cécile Huet
Funding Institution: European Commission (European Union) **Contract number:** FP7-ICT-IP-231722
Funded under: Seventh Framework Programme (2007-2013)
Work Programme Theme 3: ICT – Information and Communication Technology
Call identifier: FP7-ICT-2007-3 Challenge 2: Cognitive Systems, Interaction, Robotics
Objective: ICT-2007.2.2 Cognitive Systems, Interaction and Robotics
Start: 01/01/2009 (start of scientific work: 01/05/2009) **End:** 30/04/2013 **Total duration:** 52 months
Total EU Funding: 5.899.884 euros **Total Budget (EU Funding + Cofunding):** 7.726.783 euros

3. Project context and objectives

Vision

Year 2028, production division of the *European Advanced Robotics, Inc.*

Scenario 1. CLEVER-K26.1, a humanoid robot, fresh from the production line, is set on a chair in front of a table in a post-production "kitchen chamber". The table is spread with forks, spoons, plastic plates and glasses with various colours, pots and pans with different shapes, handkerchiefs, plastic bottles and other kitchen utensils. The edge of the table is surrounded by walls that prevent the objects from falling on the floor.

Day 0. When switched on, the robot starts slowly to move its arms, hands, head and eyes in an apparently random and uncoordinated fashion, occasionally making contact with the objects spread on the table.

Day 7. After one week of autonomous activity, the robot seems to be focussed on exploring its own body: it performs regular repetitive movements with an arm, observes its limb's displacement in space with its eyes, then focuses on its hands. It goes on to produce other repetitive movements which are a variant of the previous ones. Its "attention" seems to be always attracted by new movements which cause visual and proprioceptive unexplored/unexpected effects.

Day 30. The robot seems to have acquired basic coordination of its arms, hands and eyes, and to have started to direct attention to the outer world. Indeed, its actions are now directed to touch, press, hit, scratch, push and wipe the objects spread on the table. Suddenly an object gets accidentally caught among its fingers and is lifted in the air. This unusual event grabs the robot's attention: now it repeatedly tries to reproduce the unexpected effect. Notwithstanding its efforts, these attempts fail, so the robot directs its attention to other activities. Interestingly, a few days later, when the robot has refined its finger control, the event repeats again: this time the robot efforts succeed to reproduce the event and so the robot reproduces it several times until it masters the object-lifting action. Then its attention is directed elsewhere, in search of new events.

Day 90. After several weeks of wholly autonomous learning and interaction with the environment, the robot has acquired a variety of notable skills to interact with objects: now it 'bangs' various objects with a spoon, displaces glasses on the table by grasping them, sorts out cutlery in space, builds piles of plates, moves glasses using forks, wraps knifes in handkerchiefs.

At this point a company employee, from behind a mirror window (its presence in the chamber would immediately attract the robot's attention and disrupt its activities), starts to remotely act on the robot controller using an electronic device. The device is used to "reward" the activities that will be useful for the robot's users: putting cutlery inside the cutlery container, piling dishes, settling plates and cutlery in a precise order, etc. The reward commands are accompanied by voice commands ("settle cutlery", "pile dishes", etc.) so that in a later state the robot will be directed by voice. Indeed, CLEVER-K26.1 will be sold as an assistant robot capable of helping in house work (his curiosity drives will be eventually switched off to avoid bothering people. . .).

Scenario 2. In another part of the factory (the "play board chamber") CLEVER-B21.2, a humanoid robot with the same controller of CLEVER-K26.1 but with a smaller and more attractive physical structure is set on a chair in front of a board with various toys. The toys have various shapes, colours, and tactile textures, and posses a number of physical and electronic devices that make them to react in various surprising ways if manipulated by the robot, e.g. by emitting sounds, changing shape, becoming attached/detached from the board. CLEVER-B21.2 will undergo an experience similar to its kitchen companion but, as it will interact with a very different environment through a rather different body, it will develop very different skills: indeed, CLEVER-B21.2 will be employed as an educational robot with normal children and children with *autism syndrome*. . .

How can we create truly intelligent machines and robots? This goal has both a huge technological and scientific importance. As a technology, intelligent machines can be exploited to improve the quality of human life. Thus, they could be used to perform repetitive tasks that humans do not like to carry out, and conduct missions in hostile environments such as the ocean and deep space and those which are heavily polluted or radioactive. On the scientific side, the ability to construct truly intelligent machines can shed new light on the fundamental computational mechanisms underlying humans' learning and intelligence. Aside from the intrinsic value of this for technology in the long term, a deeper understanding of the human brain will enable a better treatment of psychiatric and neurological disorders.

Notwithstanding its importance, in the past decades scientists have failed to achieve such important goal. In particular, they have followed three main approaches to build intelligent machines and robots: (a) the construction of *knowledge-based systems* directly programmed to perform specific tasks; (b) the construction of systems endowed with *learning capabilities*, capable of acquiring specific skills on the basis of a suitable training process; (c) the development of systems on the basis artificial *evolutionary techniques* inspired by organisms' evolution. All these approaches have led to important results but have limits of *flexibility* and *scalability* (Weng et al., 2001). In particular, the main limitation of various systems and robots developed with these approaches is that they are usually designed, trained or evolved to accomplish *single specific pre-defined tasks* in only *one kind of environment* (Marshal et al., 2005). This causes them to exhibit rather limited, repetitive, fragile, and rigid behaviours, that is to be rather dull. In comparison, *real organisms* are capable of acquiring general skills that permit them to be very versatile, reactive, and adaptive with respect to different challenges and varying environmental conditions. These capabilities rely on task-independent open-ended learning mechanisms that allow them to continually adapt along the whole course of life.

Recently, a new "paradigm" (variously defined as *autonomous mental development, open-ended learning, intrinsically motivated reinforcement learning*), has been proposed directed to give robots a flexibility comparable to that of organisms (Weng et al, 2001; Di Paolo, 2002; Prince and Demiris, 2003; Blank and Meeden, 2005; Marshal et al., 2005; Blank and Meeden, 2006; Schembri et al., 2007a-c). This new paradigm stresses the need of relying on autonomous task-general learning for building intelligent robots. Several key events indicate that this new paradigm is receiving a growing attention within autonomous robotics: (a) the seminal *Workshop on Development and Learning (WDL)* (McClelland et al., 2000); (b) the construction of a website on *Autonomous Mental Development* (http://www.mentaldev.org/); (c) the new series of *ICDL - International Conference on Development and Learning* (McClelland and Pentland, 2002; http://www.egr.msu.edu/icdl02/); (d) the new series of *EpiRob - International Workshop on Epigenetic Robotics* (Balkenius et al., 2001); (e) the publication of various *Journal special issues* (*Adaptive Behavior*: Di Paolo, 2002; Prince and Demiris, 2003; *Connection Science*: Blank and Meeden, 2006; *IEEE Transaction on Evolutionary Computation*: McClelland et al., 2006).

In line with these trends, the IM-CLeVeR project aimed to develop a **new design methodology for building intelligent robots** based on intrinsically-motivated cumulative learning of skills. The central idea behind this new design methodology is that, instead of directly programming, training or evolving a set of specific skills in robots, we should endow them with *developmental programs* that allow them to autonomously develop the needed skills on the basis of *prolonged periods of interactions* with the environment and *intrinsic motivations*. They could then use the general abilities so acquired as building blocks for the solution of tasks which are relevant for the robot's users. Notice how these type of processes mark some of the most intelligent aspects of complex organisms' behaviour, in particular human and non-human primates. For example, children at play carry out several activities driven only by intrinsic motivations such as curiosity. These activities allow them to acquire knowledge and skills exploited in later adult stages to pursue useful goals. The main objectives of the project were pursued with these phenomena in mind. The scenarios presented at the beginning of this section are intended to give an overall idea of how these principles could be transferred to robots. The project used such scenarios as "guiding far landmarks" to organise the demonstrators of the proposed models.

The **central working hypothesis of the project** was that cumulative, open-ended learning in artificial systems must be based on three fundamental principles:

1. ***Hierarchical recursive architectures***. *Cumulative learning robots' control architectures should have the capability of developing sensorimotor and cognitive skills in an incremental hierarchical fashion. This requires: (a) acquiring skills and systematically increasing their complexity; (b) learning new skills using previously acquired skills as building blocks; (c) storing new skills without forgetting (and possibly improving) previously acquired ones.*
2. ***Novelty detection and intrinsic motivations***. *Cumulative learning robots need internal drives that focus learning on skills that: (a) are novel for the robots; (b) are within the robots' "zone of proximal development" (Vygotsky, 1978), that is the robots have the drive to acquire new skills which can indeed be acquired on the basis of those already acquired. To achieve this, the robots should be endowed with "intrinsic motivations" that lead them to autonomously engage in activities that produce the maximum learning rate and information gain. Internal motivations differ from "external motivations and rewards" as the latter are associated with the practical outcomes that actions produce on the external world (e.g., food or sex in organisms or accomplishment of users' goals in robots). Intrinsically motivated learning must rely on "novelty detectors", devices capable of monitoring and measuring the level of subjective novelty of actions' outcomes and learning rates so as to focus on suitable experiences and boost learning speed.*
3. ***Sensory abstraction and attention***. *Although sensory abstraction is a widely-investigated topic in cognitive sciences (e.g. in computer vision), the project will aim to isolate and study the peculiar problems of abstraction related to the specific topics of the project, namely novelty detection and hierarchical architectures for cumulative learning.*

Scientific and Technological Objectives (with relation to milestones)

Given the **knowledge gap** illustrated in the previous section, as mentioned there the overall goal of the project was the **development of a new design methodology** which permits robots to acquire increasingly complex general skills, in an open-ended fashion, thanks to intrinsically-motivated autonomous interactions with the environment. Subsequently, this allows the robots to learn to solve user-defined tasks by composing with versatility the basic skills in their repertoires. To reach this overall goal, four major scientific and technological objectives were identify (see Figure below that sketches the interdependencies of the three principles forming the project's working hypothesis and the of the four S&T challenges):

1. The main scientific challenge was to increase our knowledge of how cumulative learning is achieved in natural organisms. To do this, the project involved the implementation of ***empirical non-invasive experiments*** on intrinsically motivated learning in monkeys, children, human adults and Parkinson patients, on the basis of two novel experimental paradigms.

Milestones achieved:

- Design of novel experimental procedures for experiments with monkeys, children, adults.
- Evaluation of experimental paradigms, redesign; evaluation of first results.
- Final evaluation of results.

2. A second challenge was to provide insights regarding the mechanisms underlying intrinsically motivated cumulative learning in natural organisms. To this purpose, the project developed ***bio-mimetic models*** (including both computer simulations and robotic experiments) aiming at reproducing and explaining the empirical findings provided by the aforementioned experiments. Aside its scientific value, this effort also isolated new computational principles exploitable in robots.

Milestones achieved:

- Models capable of developing abstract features, and to guide the development of attention and vergence, on the basis of intrinsic motivations and system's goals.
- Models developing invariant visual features and interact with objects in 3D setups.
- Models of organisms' intrinsic and extrinsic motivations, models of neural plasticity underlying intrinsic dopamine-based learning in basal ganglia, models of the joystick experiment
- Models of brain mechanisms underlying both intrinsic and extrinsic motivations, models of dopamine-based action-outcome learning, models of the joystick experiment
- Models of organisms' cumulative development, and models of the hierarchical organisation of actions, tested within the board experiment setup.

3. The third main objective of the project was the development of ***new machine learning techniques***, architectures, and learning algorithms for the optimal design of cumulative learning robots. In particular, the project aimed at making substantial progress in the three distinct but related areas involved by the three principles of the project working hypothesis: (1) hierarchical architectures, (2) intrinsic motivations based on novelty detection, (3) perceptual abstraction and attention.

Milestones achieved:

- Development of algorithms capable of focusing perception in areas with change, and development of systems capable of performing predicting in observable environments.
- Development of algorithms capable of extracting dynamic, time-independent information from sensory data, and of systems performing prediction in partially observable environments.
- Development of efficient algorithms for novelty detection and attention focussing with static images, and of information-theory based novelty detectors tested in simplified kitchen scenarios.
- Development of efficient algorithms for novelty detection with dynamic images, and development of information-theory based novelty detectors tested in more challenging kitchen scenarios.
- Development of hierarchical architectures suitable to implement intrinsically-motivated cumulative reinforcement learning, prediction, planning and novelty detection.

4. The fourth main objective was to ***integrate*** the knowledge gained by the empirical experiments, the bio-mimetic computational models, and by the development of machine learning techniques and to

apply this knowledge for building real robots demonstrating cumulative learning abilities. This involved the use of a common robotic platform for the development of ***two demonstrators***, each corresponding to a simplified form of one of the two scenarios presented at the beginning: the ***kitchen scenario*** (**CLEVER-K:** more technological oriented) and the ***board experiment scenario*** (**CLEVER-B:** this models the "board experiments" carried out in the project with monkeys and children).

Milestones achieved:

- First design of the robotic bio-mimetic demonstrator's integrated architecture.
- First design of the machine-learning demonstrator's integrated architecture.
- Robotic bio-mimetic demonstrator that performs explorations and acquires skills by interacting with the board and objects of the board experiment setup.
- Robotic machine-learning demonstrator that is capable of autonomously developing an action repertoire, on the basis of intrinsic motivations, in the kitchen scenario.
- Robotic bio-mimetic demonstrator capable of acquiring the value of events on the basis of external rewards in the board experiment.
- Robotic machine-learning demonstrator capable of autonomously developing compound skills on the basis of pre-existing skills and intrinsic motivations.
- Robotic bio-mimetic demonstrator capable of exploiting knowledge, learned on the basis of intrinsic motivations, to pursue goals which acquired value with external rewards.
- Robotic machine-learning demonstrator capable of using previously acquired skills (e.g. by assembling them in sequence or hierarchy) to accomplish externally rewarded goals.

The accomplishment of the four objectives was fueled by a theoretical analysis organised around the compilation of a Roadmap Book, another milestone of the project. The activities culminated into two International Workshops that established the field's state of the art, two summer schools, and a final Research Topic of Frontiers in Cognitive Science/Neurorobotics which collected the major results and recommendations produced by the project.

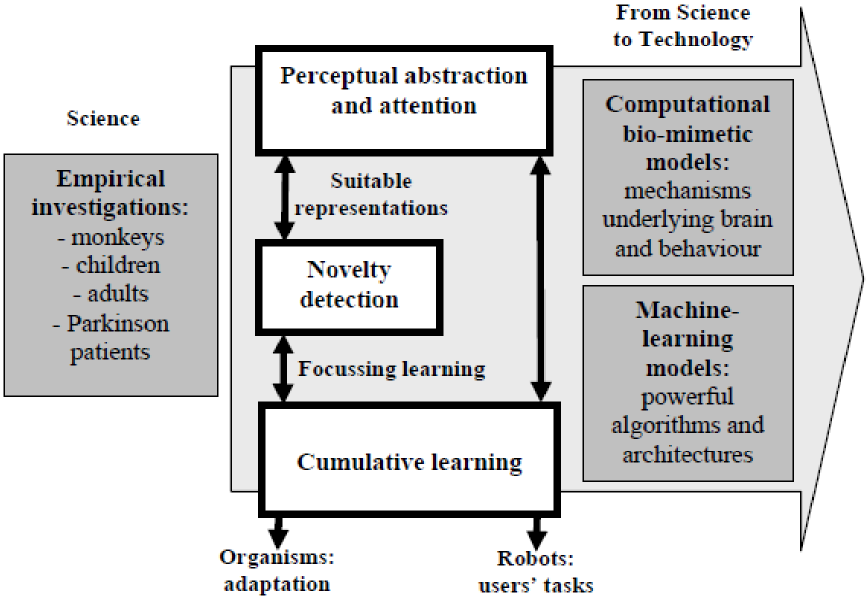
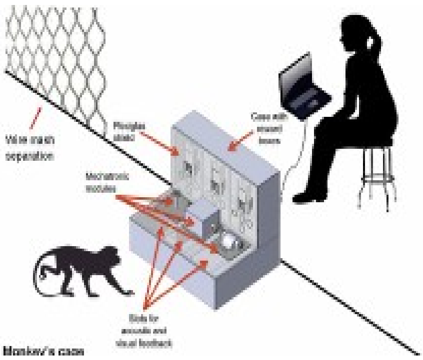


Figure. The working hypothesis of the project and its main scientific and technological challenges. *White bold boxes and vertical arrows* indicate respectively the three pillars of the working hypothesis of the project and their logical interdependencies. *Gray boxes and horizontal arrow* indicate the logical implementation of the project, from the scientific challenges (*left*) to the technological challenges (*right*), passing through the challenges related to the project's bio-mimetic and machine learning modelling (*centre*).

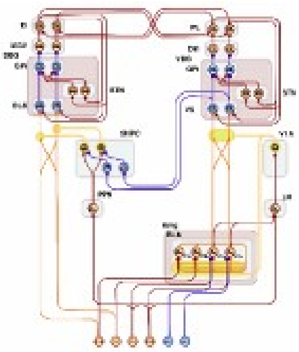
4. Overview of work performed and main results achieved

The project achieved important results in various directions: it defined innovative experimental paradigms to investigate intrinsically-motivated action learning, it achieved an important theoretical understanding of intrinsic motivations in organisms and robots, it developed bio-constrained models of intrinsic motivations and hierarchical behaviour, it devised new machine learning algorithms for solving complex sensorimotor problems, and finally it implemented and tested integrated architectures to investigate all these issues within the three iCub robots available at the Teams' labs. More in detail, the most important achievements of the project were as follows.



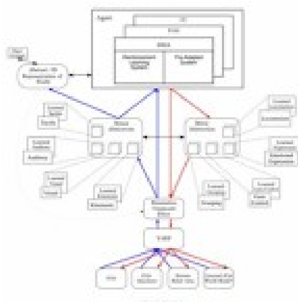
Mechatronic board and empirical experiments.

- Development of a new experimental paradigm to investigate learning of action-outcomes and actions performed on a joystick, usable with rats and humans
- Construction of a novel flexible Mechatronic Board usable to investigate intrinsically-motivated learning in monkeys, children, human adults, and humanoid robots
- Formulation of new experimental protocols to investigate intrinsically motivated cumulative learning in monkeys, children, and human adults.



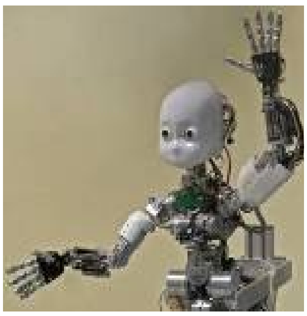
Development of bio-constrained computational models

- Bio-constrained models of extrinsic motivations, and of intrinsic motivations with predictors and habituators regulating dopamine
- Theory and robotic models of the staged development of sensorimotor skills in children
- Models of the brain dopamine system and basal ganglia mechanisms at the basis of intrinsically motivated learning of actions
- Models of brain systems at the basis of intrinsically motivated cumulative learning: basal ganglia-cortical loops and cortico-cortical pathways implementing habits and goal-directed behaviour



Development of machine learning models

- New algorithms for visual perception and abstraction
- Bio-inspired methods for novelty detection based on habituation
- Information-theory based approaches to intrinsic motivations
- New hierarchical reinforcement learning controllers
- Innovative predictive models of robots' body and environment
- Sofware and hardware infrastructures for the implementation of integrated robotic demonstrators



The integration work and the robotic demonstrators

- A machine-learning robot controller capable of: (a) developing new actions on the basis of intrinsic motivations (b) re-using and modifying them to solve new tasks
- A bio-constrained integrated architecture capable of (a) learning actions/action-outcomes based on intrinsic motivations (b) using goals to recall actions with extrinsic motivations

5. Overview of main results per partner

ISTC-CNR-LOCEN

- Contributed to define the mechatronic board and the experimental protocols with monkeys and children: this is now a novel unique experimental tool that can be used to investigate intrinsically motivated cumulative learning in monkeys, children and robots.
- Played a key role in developing a general theory on intrinsic and extrinsic motivations. Clarified the distinction between prediction-based, novelty-based, and competence-based intrinsic motivations. Developed several bio-constrained models of cumulative learning based on these different types of intrinsic motivations.
- Highlighted theoretically, and based on models, how the cumulative learning of skills (competence) based on intrinsic motivations needs to pivot on action *goals* and action-outcome representations.
- Developed general theories and system-level bio-constrained models of the hierarchical organisation of brain: these theories and models allows capturing a number of behavioural phenomena related to autonomous cumulative learning, goal-directed behaviour, and habits formation and exploitation.
- Developed an integrative bio-constrained model of extrinsic motivations (based on amygdala) and of the brain architecture underlying hierarchical sensorimotor behaviour, based on the main striato-cortical loops (limbic, associative, sensorimotor).
- Developed new bio-inspired hierarchical-reinforcement learning models that can solve multiple tasks by suitably allocating "expert modules" to them based on the sensorimotor complexity of the tasks. The model works with continuous states and actions in the iCub robot.
- Coordinated and played a major role in the design and implementation of CLEVER-B demonstrators involving CNR, AU and USFD. This is the first bio-constrained robotic model that can learn action-outcome associations based on intrinsic motivations, an can recall them based on the internal reactivation of goals by extrinsic motivations.

ISTC-CNR-UCP

- Detailed an experimental protocol, experimental set-up to investigate intrinsic motivations in monkeys. Participated to define the features of the mechatronic board.
- Ran two sets of experiments with monkeys and the mechatronic board, collected and analysed the data and, on this basis, showed how intrinsic motivations can guide learning in monkeys.

ISTC-CNR-Barto

- Fostered theoretical and modelling ideas on intrinsic motivations within the whole project, with visits and participations to various activities of the project.
- Developed state-of-the-art hierarchical reinforcement learning and intrinsic motivation system.

UCBM

- Designed and implemented the mechatronic board and the experimental protocols related to it. The board is a new, unique experimental tool that can be used to investigate learning based on intrinsic motivations with primates and robots.
- Run the experiments with children based on the mechatronic board, collected and analysed the data, and, on this basis, showed how children learning can be informed by intrinsic motivations.

USFD

- Developed the experimental paradigm `Joystick Task' enabling the study of intrinsic motivations in humans and animals.

- With colleagues in Japan, developed a variant of the `Joystick Task' using monkeys and saccadic eye movements, allowing us to more directly study brain mechanisms involved with action acquisition and already producing a wealth of data (e.g., signals originating exclusively from subcortical visual processing is sufficient for development of novel actions).
- Completed experimental study with humans showing that action acquisition using visual reinforcement signals that are not directly-available to subcortical structures is impaired, supporting neuroscientific theories underlying the IM-CLeVeR project.
- Developed biologically-inspired neural network model of the `Joystick Task' that describes `intelligent' exploration strategies using only simple mechanisms of the basal ganglia. Several testable predictions are being tested with experimental studies.
- Developed a model showing how a period of stable behaviour can arise from a learning rule that does not incorporate any notion of optimality; such behaviour diverges with overtraining, providing a potential explanation for why dopamine signal must habituate in novelty-based action discovery.

FIAS

- Implemented on iCub a bio-inspired general-purpose vision system capable of autonomously exploring the environment and determining the subset of relevant objects that will be subsequently learnt for future recognition based on intrinsic motivations.
- Developed new biologically plausible systems for vergence control and eye-head coordination learning based on intrinsic motivations.

UU

- Successfully developed a new approach for segmentation of objects using live streams from robot cameras in natural scenes using a 3D approach. Successfully implemented a Bayesian based method for hierarchical representation of data and demonstrated that the approach is an effective method for information storage for a hierarchy of images.
- Developed an online method of detecting features and recognising objects for the novelty detector. Developed a new novelty detector that addresses limitations of previous ones and was validated in experiments with physical robots in real world environments.
- Implemented techniques to enable a physical robot to carry out actions on perceived objects and identify the outcome of these actions so that basic affordances of the objects can be associated with particular events. In particular, developed an evolutionary approach for robots to create new skills based on an appropriate combinations and sequencing of lower lever skills.

AU

- Conducted an extensive literature review of infant psychological and neurological development from conception to 12 months postnatal, which is underpinning work on staged development in the iCub.
- Implemented a number of models for staged eye and head saccade learning, and constrained reaching, on the iCub.
- Implemented robotic systems whose development is based on different types of constraints, that it contributed to classify, and intrinsic motivations.
- Designed and implemented several interfaces between computational models used by different partners in relation to CLEVE-B models.

IDSIA-SUPSI

- Developed a humanoid planning framework to create task-relevant roadmaps, which can be used to perform smooth motions and build a basis for learning task-relevant behaviours.
- Extended the capabilities of icVision, a computer vision and hand-eye coordination framework, which allows in combination with CGP-IP, our Cartesian Genetic Programming implementation, the learning

of visual representations of objects in the scene. These allow detection, identification and localisation in real-time, which is an important requirement for achieving manipulation. In particular, thanks to a collaboration with FIAS, we can autonomously explore the scene and learn our representations without the need of a human.

- Developed MoBeE, a modular behavioral environment for humanoids and other robots, which integrates elements from vision, planning, and control, in order to facilitate the synthesis of autonomous, adaptive behaviors.
- Developed Modular-Least Squares Policy Iteration (M-LSPI), a novel method that enables real-world application of LSPI to massive Markovian reinforcement learning problems through modular/hierarchical decomposition.
- Created the Upper Confidence Weighted Learning (UCWL) framework for calculating intrinsic rewards through estimating the confidence intervals of the agent's predictions, which allows for efficient exploration in human-robot interaction scenarios with incomplete feedback.
- Introduced Curiosity-Driven Modular Incremental Slow Feature Analysis (CD-MISFA), a hierarchical curiosity-driven learning system that autonomously learns multiple abstract slow-feature based representations from a robot's high-dimensional visual input stream.
- Introduced PowerPlay, a way of automatically inventing the simplest still unsovable problems. Conducted first successful PowerPlay experiments based on recurrent neural networks.
- Based on successful collaboration between the project partners, IDSIA and FIAS integrated various vision-based methods in the CLEVER-K demonstrator and published their results together.

6. Main achievements in detail, per partner

CNR-ISTC-LOCEN

The Team played a key role in developing a general theory on intrinsic and extrinsic motivations.
During the four years of the project, the Team, also supported by an intense collaboration with Prof. Barto, gave an important thrust to the research of the project on the theoretical understanding and modeling of intrinsic motivations and their difference with respect to extrinsic motivations. A good part of this work has fueled the "Intrinsic Motivation Book" (Baldassarre and Mirolli, 2013a) used as a collector of various theoretical ideas developed within the project. This research has in particular supported the development of a view on IMs function for adaptation (or for solving robotic tasks): IMs are cognitive components *driving the acquisition of knowledge and skills in the absence of learning signals relying on extrinsic rewards that can directly improve fitness* (e.g., finding food or water, in the case of organisms; or solving the "user tasks", in the case of robots) (Baldassarre, 2011). When these extrinsic rewards are absent, then learning signals can be generated by IMs on the basis of the detection of the rates of learning of other components of the brain/controller. In a later stage, the knowledge and skills so acquired can be used to accomplish the extrinsic rewards when this becomes possible so to improve fitness (or, in robots, to accomplish the user's tasks).
We also identified *3 main types of IM mechanisms plus a 4th "special case"*:
(1) *Knowledge-based IMs: prediction-based IMs*. These are IMs related to the acquisition of "diachronic knowledge", that is on the prediction of future states based on current states and (possibly) planned actions. In general, these IMs rely on the knowledge-level, or learning rate, of forward models, e.g. generate learning signals (or the motivations to drive behaviour) when the forward model to which they refer makes substantial errors or is learning to decrease them in a fast way (Baldassarre and Mirolli, 2013a; Barto et al., 2013).
(2) *Knowledge-based IMs: novelty-based IMs*. These are IMs related to the acquisition of "synchronic knowledge", that is on the sensory appearance and inner structure of world states in a given instant. In particular, these IMs are based on the novelty of experienced objects, object-object combinations, and context-object combinations. These IMs require a memory (knowledge) of previous experiences and are based on the fact that the current experience is similar or novel with respect to them (Baldassarre and Mirolli, 2013a; Barto et al., 2013).
(3) *Competence-based IMs*. These are IMs based on a measure of the acquisition of competence, that is of the skills (i.e., action policies) that allow the agent to change the world as desired. These IMs lead an agent to

engage in pursuing those goals for which the learning rate of the skills (competence), or inverse models needed to accomplish them, is high (Miroli and Baldassarre, 2013).
(4) In the late years of the project, the Team has also contributed to understand the importance of the *problem of the autonomous generation of goals based on IMs for having truly open-ended cumulative learning systems:* this is one of the most important insights of IM-CLeVeR and a possible important starting point for future research (Baldassarre et al., 2012; Santucci et al., in preparation).

The Team developed various models showing the importance of bottom-up and top-down attention, and on the coupling between attention and pragmatic (e.g., reaching) actions.
The Team modeled bottom-up and top-down attention showing how when used together bottom up and top-down attention can suitably integrate and lead to boast the learning processes of top-down attention. The same works showed the importance of having models with a strong coupling between attention and pragmatic actions: this lead has a great importance as “action” is so decomposed into its “where” component (attention) and “what” component (e.g., what the arm and hand do; Ognibene et al., 2010, Ognibene et al., submitted).

The Team has developed an effective methodology for building system-level bio-constrained computational models.
The research of the Team on the hierarchical organisation of brain and behaviour is based on a system-level modelling approach called *Computational Embodied Neuroscience (CEN)* (Caligiore et al., 2010; see Baldassarre, 2012, for the best example of the application of the method). The idea of CEN is to build models that are cumulative, i.e. they can account for an increasing number of experimental data, rather than to build one-use-only models to account for one specific phenomenon and then to start a new one to account for a new phenomenon (as often happens in the modeling literature). To this purpose, CEN builds models that ideally satisfy four constraints: (a) the macro-architecture of the model, and possibly its micro-architecture and functioning, are constrained with neuroscientific data; (b) the behaviour of the model is constrained with behavioural data, and the model is requested to account for a progressively increasing number of behaviours; (c) the model should reproduce and account not only for the expression of behaviour, but also the learning processes that lead to its acquisition; (d) the model should be sufficient to autonomously work within an embodied and situated set-up. Importantly, the request of cumulation and embodiment naturally leads to build system-level models capturing whole sub-systems of brain capable of implementing the multiple functions needed to actually interact with the environment.
The approach also has three critical advantages: (a) It leads to an important overall understanding of brain that is often missing in the neuroscientific and psychological literature: this allows the Team to easily give important theoretical and modeling contributions to several empirical investigations. (b) It is an ideal means to transfer knowledge from neuroscience/psychology to robotics as the models produced are already capable of controlling embodied systems such as robots. (c) Last, and probably most important, the approach produces *cumulation of knowledge*: and cumulativeness is the hallmark of science.

Developed general theories and system-level bio-constrained models of the hierarchical organisation of brain: these theories and models allow capturing a number of behavioural phenomena related to autonomous cumulative learning, goal-directed behaviour, and habits formation and exploitation.
During the fours years of the project, the application of the CEN research method to understand the hierarchical organisation of brain and behaviour has revealed that hierarchy in brain can be found, and investigated, at three levels:
(1) Sub-cortical hierarchies. During the project, the Team has been working and publishing various models describing at an increasing level of sophistication and detail the sub-cortical organisation of brain. We are now working on a journal article (Mannella et al., in preparation) that summarises this work on the basis of a system-level model. This model views the sub-cortical organisation of brain as pivoting on three main basal ganglia-cortical loops: a sensorimotor loop selecting actions; an associative loop supporting attention and action programming; and a limbic loop supporting the formation and selection of goals.
(2) Cortical hierarchies. The hierarchical organisation of behaviour is also reflected in the hierarchical organisation of cortex. The starting point of our research on this was the important article Caligiore et al. (2010). This paper proposed model, named TRoPICALS, that incorporates a functional/anatomical hypothesis on the macro architecture of hierarchical brain cortex underlying executive control and sensorimotor behaviour. The strength of the model is that its architecture and functioning integrates three key principles: (a)

the overall organisation of cortical brain into the dorsal and ventral neural pathways; (b) the processes that support affordance detection and action selection within the dorsal pathway; (c) the processes that support top-down executive control on affordances and actions within the dorsal pathway, and the influence on this control by limbic/affective processes.

(3) Integrating sub-cortical and cortical hierarchies. In the four years of the project, the investigation of brain hierarchy has led us to understand that sub-cortical and cortical hierarchies are usually studied in isolation by different neuroscience sub-communities. Instead, we have understood that cortical and sub-cortical systems are intimately and closely related. We have so been developing an integrated view of sub-cortical and cortical hierarchies: a first presentation of this view is given in Baldassarre et al. (2013) but it probably deserves further elaboration and publications. This integrative approach has been already very productive as it has allowed us to give a reinterpretation of the mirror neurons system as heavily involving sub-cortical circuits aside cortical ones, as shown in the relevant publication of Caligiore et al. (accepted).

The Team highlighted theoretically, and based on models, how the cumulative learning of skills (competence) based on intrinsic motivations needs to pivot on goals and action-outcome representations.

The view on the hierarchical organisation of brain has been enriched with the elaboration of the critical concept of *goal*, that has revealed more and more important for intrinsic motivations and autonomous cumulative learning in the last two years of IM-CLeVeR (Baldassarre et al. 2012; Santicci et al., in preparation). The elaboration of the role of high-level goals for cumulative learning is probably one of the most important achievements of IM-CLeVeR. A goal is a representation of a possible outcome of actions that can be re-activated internally (e.g., based on extrinsic or intrinsic motivations) and on this basis is capable of recalling the action(s) that lead to its accomplishment. The theoretical journal paper of Mannella et al. (accepted), done in collaboration with USFD, is sharpening these ideas by theoretically investigating the specific mechanisms through which extrinsic and intrinsic (novelty based) motivations support the attribution of value to goals and, on this basis, their selection. These insights on goals, and the previous work with TRoPICALS (Caligiore et al., 2010), has led to a theoretical article published in a prestigious neuroscientific journal led by us (Thill et al., 2013).

The Team developed new hierarchical-reinforcement learning models that can solve multiple tasks by suitably allocating "expert modules" to them based on the sensorimotor complexity of the tasks: the model was tested in the iCub robot and is capable of implementing transfer learning, a fundamental process for cumulative learning.

The Team has developed hierarchical reinforcement-learning systems capable of implementing transfer of knowledge between skills while learning multiple tasks in sequence. An important insight has been that transfer is necessary for efficient and fast cumulative learning when several tasks have to be solved, a conditioned required by future autonomous robots. In particular, the system developed here is able to: (a) encode in the same neural structures ("expert modules") similar skills so as to favor generalization and fast learning; (b) encode in different neural structures different skills so as to avoid catastrophic interference. The system, that can function with continuous states and actions, has been shown to scale up to control the iCub arm working with 4 DOFs. The system can be also used for developmental psychology modeling, in particular in relation to the *assimilation* and *accommodation* processes first introduced by Piaget. This research has produced conference publications (Caligiore et al., 2010; Tommasino et al., 2012a) and a Journal article (Tommasino et al., in preparation).

The Team coordinated and played a major role in the design and implementation of CLEVER-B Demonstrators involving CNR, AU, USFD, and FIAS: CLEVER-B models are the first bio-constrained robotic models that can learn action-outcome associations based on intrinsic motivations, an can recall them based on the internal re-activation of goals by extrinsic motivations.

CLEVER-B Demonstrators aimed to build robot controllers capable of undergoing an intrinsically motivated cumulative learning based on biologically constrained mechanisms. The robots shown in the Demonstrators integrated the psychologically-constrained sensorimotor mechanisms and actions of AU, the biologically-constrained decision-making components of CNR-ISTC-LOCEN and USFD, and the biologically constrained visual abstraction and attention mechanisms of FIAS. The Demonstrator aimed to: (a) show that intrinsic motivations (IM) are critical for learning action-outcome contingencies ("agency") later re-usable to achieve extrinsic rewards and goals; (b) understand the neural mechanisms supporting IM in animals; (c) transfer the

knowledge on these issues to robots. The key ingredients of the architecture are as follows:
- Bottom-up and top-down attention system (based on Ognibene et al., submitted).
- Novelty-based IM to guide exploration; prediction-based IM to guide the formation of inverse models (Baldassarre and Mirolli, 2013a; Barto et al., in preparation).
- Competence-based intrinsic motivations to learn the actions (Mirolli and Baldassarre, 2013; Mirolli et al., 2013; Santucci et al., in preparation).
- Hierarchical sub-cortical/cortical organisation of brain to store the acquired knowledge and skills (Baldassarre and Mirolli, 2013; Thill et all, 2013; Mannella et al., accepted; Caligiore et al., 2013).
The first three Demonstrators have led to three journal papers: CLEVER-B1: Chersi et al., 2012; CLEVER-B2: Baldassarre et al., 2012; CLEVER-B3: Fiore et al., in preparation.

Selected bibliography

Baldassare, G. (2011). What Are Intrinsic Motivations? A Biological Perspective. In Cangelosi, A.; Triesch, J.; Fasel, I.; Rohlfing, K.; Nori, F.; Oudeyer, P.-Y.; Schlesinger, M. & Nagai, Y. (ed.), Proceedings of the International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob-2011). E1-8.

Baldassare, G.; Mannella, F.; Fiore, V. G.; Redgrave, P.; Gurney, K. & Mirolli, M. (2012), 'Intrinsically motivated action-outcome learning and goal-based action recall: A system-level bio-constrained computational model', *Neural Networks,* 41, 168-187.

Baldassarre, G. & Mirolli, M., ed. (2013), *Computational and Robotic Models of the Hierarchical Organisation of Behaviour*, Springer-Verlag, Berlin.

Baldassarre, G. & Mirolli, M., ed. (2013a), *Intrinsically Motivated Learning in Natural and Artificial Systems*, Springer-Verlag, Berlin.

Baldassarre, G.; Caligiore, D. & Mannella, F. (2013), The hierarchical organisation of cortical and basal-ganglia systems: a computationally-informed review and integrated hypothesis, *in* Gianluca Baldassarre & Marco Mirolli, ed., 'Computational and Robotic Models of the Hierarchical Organisation of Behaviour', Springer-Verlag, Berlin.

Barto, A., Mirolli, M., Baldassarre, G. (in preparation). Novelty or suprise? *Frontiers in Cogntive Science.*

Caligiore, D., Borghi, A., Parisi, D., Baldassarre, G. (2010). TRoPICALS: A Computational Embodied Neuroscience Model of Compatibility Effects. *Psycological Review*, 117 (4), 1188-1228.

Caligiore, D., Mirolli, M., Parisi, D., & Baldassarre, G. (2010). A bioinspired hierarchical reinforcement learning architecture for modeling learning of multiple skills with continuous state and actions. In: *Proceedings of the Tenth International Conference on Epigenetic Robotics (EpiRob2010)*. pp. 27 - 34. Johansson B., Sahin E. and Balkenius C (eds.). (Lund University Cognitive Studies, vol. 149). Lund, Sweden: Lund University.

Caligiore, D.; Pezzulo, G.; Miall, C. R. & Baldassarre, G. (accepted), 'The contribution of brain sub-cortical loops in the expression and acquisition of action understanding abilities', *Neuroscience and Biobehavioral Reviews*.

Chersi, F.; Mirolli, M.; Pezzulo, G. & Baldassarre, G. (2012), 'A spiking neuron model of the cortico-basal ganglia circuits for goal-directed and habitual action learning', *Neural Networks,* 41, 212-224.

Fiore, V. G.; Sperati, V.; Mannella, F.; Dolan, R. J. & Baldassarre, G. (in preparation), 'Keep focussing: striatal dopamine multiple functions resolved in a single mechanism tested on the iCub', *Frontiers in Cognitive Science*.

Mannella, F.; Gurney, K. & Baldassarre, G. (accepted), 'The nucleus accumbens as a nexus between values and goals in goal-directed behaviour: a review and a new hypothesis', *Frontiers in Behavioural Neuroscience*.

Mannella, F.; Mirolli, M. & Baldassarre, G. (in preparation), 'Instrumental devaluation may depend on Pavlovian associations between manipulanda and rewards: a computational model', *Frontiers in Neuroscience*.

Mirolli, M. & Baldassarre, G. (2013), Functions and mechanisms of intrinsic motivations: the knowledge versus competence distinction, *in* Gianluca Baldassarre & Marco Mirolli, ed., 'Intrinsically Motivated Learning in Natural and Artificial Systems', Springer-Verlag, Berlin, pp. 49--72.

Mirolli, M.; Baldassarre, G. & Santucci, V. G. (2013). Phasic dopamine as a prediction error of intrinsic and extrinsic reinforcement driving both action acquisition and reward maximization: A simulated robotic study. *Neural Networks*, 39 40-51.

Ognibene, D.; Pezzulo, G. & Baldassarre, G. (2010). How can bottom-up information shape learning of top-down attention-control skills?. In Kuipers, B.; Shultz, T.; Stoytchev, A. & Yu, C. (ed.), Proceedings of 9th IEEE International Conference on Development and Learning (ICDL2010). 231-237.

Ognibene, D.; Pezzulo, G. & Baldassare, G. (submitted), 'Learning and Development of Representations and Attention for Action: A Neurorobotic Study', *IEEE Transactions on Autonomous Mental Development*.Baldassarre, G.; Caligiore, D. & Mannella, F. (2013), The hierarchical organisation of cortical and basal-ganglia systems: a computationally-informed review and integrated hypothesis, *in* Gianluca Baldassarre & Marco Mirolli, ed., 'Computational and Robotic Models of the Hierarchical Organisation of Behaviour', Springer-Verlag, Berlin.

Santucci, V. G.; Baldassarre, G. & Mirolli, M. (in preparation), 'Learning multiple skills through competence-based intrinsic motivations: a computational embodied model', *Frontiers in Neurorobotics*.

Thill, S.; Caligiore, D.; Borghi, A. M.; Ziemke, T. & Baldassarre, G. (2013), 'Theories and computational models of affordance and mirror systems: An integrative review', *Neuroscience and Biobehavioral Reviews* 37, 491-521.

Tommasino, P.; Caligiore, D.; Mirolli, M. & Baldassare, G. (2012), Reinforcement learning algorithms that assimilate and accommodate skills with multiple tasks, *in* 'Proceedings of IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob), 2012', IEEE, , pp. 1-8.

Tommasino, P.; Caligiore, D.; Mirolli, M. & Baldassarre, G. (in preparation), 'Transfer expert reinforcement learning (TERL): a reinforcement learning architecture that transfers knowledge between skills', *Neural Networks*.

CNR-ISTC-UCP

Experiments on intrinsically-motivated learning in monkeys

Animals (including humans) act as if they are endowed with complex motivational systems that drive them to do so. Intrinsic motivation, described as the drive that leads exploratory actions "for their own sake" (Hughes, 1997), is an important mechanisms underlying learning processes as it seems related to the acquisition of knowledge/skills that may be recalled and used in a later stage. Previous studies on nonhuman primates have brought circumstantial evidence that intrinsic motivation promotes exploration and learning (see Harlow, 1950, Harlow et al., 1950, Welker, 1956). In collaboration with CNR-ISTC-LOCEN and UCBM-LBRB we developed and refined during the course of the project a mechatronic board suitable for comparative investigations on (human and non-human) primates and robots (Taffoni et al., 2012; 2013). Our experiments with the mechatronic board were designed in collaboration with CNR-ISTC-LOCEN and UCBM to understand whether exploratory actions (not instrumental in achieving particular goals other than performing the actions themselves) affect the performance in subsequent problem solving tasks that require the proficiency acquired during previous spontaneous exploration. We developed and conducted two generations of experiments: the first generation allowed us to improve and refine the experimental protocol and the second generation allowed us to show the role of action-outcome contingencies in promoting learning, integrating our results with those on children and providing specific inputs to the developers in their work with robots (Polizzi di Sorrentino et al., in preparation; Taffoni et al., 2012; 2013). Throughout the project intense discussions and *ad-hoc* meetings with other teams, and in particular with CNR-ISTC-LOCEN, UCBM and USFD, were fundamental to improve the experimental design and protocol. The intense collaboration between the different teams involved in WP3 allowed converging to a common experimental setup, which led to the successful comparison of intrinsically-motivated learning processes in monkeys and children.

The empirical results on monkeys suggest that, in the absence of extrinsic reinforcement, the opportunity to discover action-outcome contingencies promotes individuals' exploratory drives and learning (Polizzi di Sorrentino et al., in preparation). When contingencies cannot be controlled by an individual's own actions, as it is the case for yoked subjects (who did not experience action-outcome contingencies), there is an evident loss of interest that prevents acquisition of competence. It is possible that the yoked subjects that passively witnessed the outcomes of the mechatronic board without producing any action experienced a "learned helplessness effect" which was strongly detrimental for learning. The results successfully evidenced the role of IM in driving versatile learning in monkeys and were discussed in light of commonalities and differences with the results on children.

From a neuroscientific perspective, the rapid decrease in exploration shown by yoked subjects closely matches what expected at the brain level. In fact, in the absence of behaviourally rewarding consequences, the phasic DA response toward unpredicted novel neutral stimuli diminishes rapidly due to habituation. Under this condition subjects are prevented from further exploring the board. By contrast, when outcomes are contingent with actions (like in the experimental condition) they may function as primary rewards, reinforcing action repetition and thus learning. So far, still little is known on how action-outcome contingencies may effectively block the decrease of phasic DA response. This research question could be further addressed by combining the neuroscientific approach (i.e., through non-invasive EEG recording of visual evoked potentials in humans) with behavioural observations on monkeys and children. Our study illustrates the importance of combining neuroscientific results with behavioural ones, and might possibly help to address future studies in the neuroscientific field.

Capuchin monkeys are very explorative and manipulative. This is also related to their way to approach things and food, which led them to be called extractive and destructive foragers. They are also able to explore and appreciate the properties of objects and combine different objects. This vigorous approach to things fosters their success in exploiting many food resources, not accessible to other species. Studies in our field site in Piaui, Brazil, are helping us to understand what kind of information, knowledge and skills animals can achieve through intrinsic motivation from the environment and what can they learn from both the physical and the social domain. Our observations and field experiments provide a way to think about how naïve individual monkeys become proficient foragers. We adopt an ecological approach which considers the phenomenon of skill acquisition in the natural contexts in which young individuals live, describes what they do that might affect their own learning, and also how they might learn in relation to the behaviour of more skilled group members (Visalberghi & Fragaszy, 2012). Moreover, the studies carried out by the Ethocebus team on tool-using skills of wild capuchin monkeys suggest that curiosity-driven exploration contributes to the acquisition of tool use (Visalberghi & Fragaszy, 2013).

Experiments on cumulative learning in monkeys

Cumulative learning processes in tool use

It has been shown that capuchin monkeys are able to efficiently detect objects' properties (i.e., when choosing a stone as tool to crack open nuts) on the basis of visual (size) and non-visual functional clues (friability, weight). In a recent work (Manrique et al., 2011), we tested capuchin monkeys in a task that required them to select a specific tool on the basis of its properties (i.e., rigidity). Subjects were allowed to infer object's property by observing an experimenter playing with the tool or by directly manipulating it. When facing an out-of-reach reward capuchins efficiently used information previously gathered about tool affordances in the absence of an extrinsic reward. The intrinsic motivation underlying free exploration may thus promote individuals' abilities to solve subsequent problem solving tasks.

Given the above results, we took advantage of this experimental set up to conduct a following experiment on cumulative learning in capuchin monkeys by exploiting their ability to use tools. Specifically, we aimed to assess whether capuchin monkeys can learn to use a tool to obtain a second tool that could then be used to obtain a final goal. Tools could differ for their rigidity/flexibility (a non-visual functional property) and capuchins had to solve two different tasks in order to use tools sequentially (i.e., use a rigid tool to retrieve a second flexible tool and then use the latter to obtain an out-of reach liquid reward). Ten capuchins were presented to four conditions (2x2 design) in which the experimenter showed rigidity properties of one rigid/flexible tool (bending and unbending it), placed it out of reach, and then handed the other rigid/flexible tool to the subject. Capuchins used the tools in sequence to get the reward when necessary (nine out of ten subjects from the first trial) and not as a result of a simple heuristic of using the tool placed nearest to the food. When the demonstrated tool was unnecessary, capuchins used the flexible tool already within reach. The current study demonstrated that capuchin monkeys are able to put discrete responses in the hierarchically correct sequence to obtain a goal using tools differing in their rigidity, a non-visually detectable property (Sabbatini et al., in preparation).

Cumulative learning processes in concept learning

Previous studies demonstrated that capuchins are able to acquire abstract concepts on the basis of perceptual equivalence between stimuli. In a previous study the ability of capuchin monkeys to learn same and different concepts was investigated by testing whether subjects use them to solve a relational matching-to-sample task (Truppa et al., 2010). The relational MTS (RMTS) task involves requires subjects to understand whether the relationship among attributes of objects belonging to one set is equivalent to the relationship among objects belonging to another set (e.g., sets of objects of the same shape), with objects belonging to different sets that have different shape. We found that only one subject, Roberta, solved the task and that, to gain the solution, she spontaneously decomposed the task in two sub-problems. First she focused on the *same* trials and reached criterion in this condition and only then she focused on the *different* condition and reached criterion in the latter condition, while her performance on the *same* condition worsened. Eventually, in the last part of the learning process, she recombined the knowledge previously acquired separately becoming concurrently successful in both conditions. On this basis we argue that a very demanding aspect, in terms of attentive resources and/or working memory load, of the same/different relational matching is to learn two concepts (sameness and difference) “at once”, that is when trials presenting *same* or *different* conditions are randomly alternated. This is the first evidence of same/different relational matching-to-sample abilities in a New World monkey and we demonstrated that the ability to match novel stimuli is within the capacity of this species (Truppa et al., 2011).

The results of our experiments were presented at various scientific conferences and published in international peer-review journals. Moreover, these results were disseminated for the general public *via* newspapers, radio and internet.

Selected bibliography

Manrique, H.M., Sabbatini, G., Call, J., and Visalberghi, E. (2011) Tool choice on the basis of rigidity in capuchin monkeys. *Animal Cognition* 14:775-786 doi: 10.1007/s10071-011-0410-9.

Polizzi di Sorrentino, E., Baldassarre, G., Bordonali, A., Formica, D., Guglielmelli, E., Sabbatini, G., Taffoni, F., Truppa, V., and Visalberghi, E. Exploration and learning in capuchin monkeys (*Cebus apella*): the role of action-outcome contingencies. *In preparation for Animal Cognition.*

Sabbatini, G., Manrique, H.M., De Bortoli, Vizioli. A., Trapanese, C., Call, J. and Visalberghi, E. Sequential use of rigid and pliable tools in capuchin monkeys (*Cebus apella). In preparation for Animal Cognition.*

Taffoni, F., Vespignani, M., Formica, D., Cavallo, G., Polizzi di Sorrentino, E., Sabbatini, G., Truppa, V., Visalberghi, E., Keller, F., and Guglielminelli, E. (2012). A mechatronic platform for behavioral analysis of nonhuman primates. *Journal of Integrative Neuroscience, 11*, 87-101.

Taffoni, F., Formica, D., Schiavone, G., Scrocia, M., Tomassetti, A., Polizzi di Sorrentino, E., Sabbatini, G., Truppa, V., Mannella, F., Fiore, V., Mirolli, M., Baldassarre, G., Visalberghi, E., Keller, F., and Guglielmelli, E. (2013). The "Mechatronic Board": a tool to study intrisic motivation in humans, monkeys, and humanoid robots, in Baldassarre, G.; Mirolli, M (Eds.) *Intrinsically Motivated Learning in Natural and Artificial Systems,* pp 411-432. Springer Berlin Heidelberg.

Truppa, V., Piano Mortari, E., Garofoli, D., Privitera, S., and Visalberghi, E. (2010). Identity concept learning in matching-to-sample tasks by tufted capuchin monkeys (*Cebus apella*). *Animal Cognition,* 13, 835–848.
Truppa, V., Piano Mortari, E., Garofoli, D., Privitera, S., and Visalberghi, E. (2011). Same/different concept learning by capuchin monkeys in matching-to-sample tasks. *PloS One* 6(8): e23809. doi:10.1371/journal.pone.0023809.
Visalberghi, E., and Fragaszy, D. (2012). Learning How to Forage: Socially Biased Individual Learning and "Niche Construction" in Wild Capuchin Monkeys. In "The Primate Mind" (Edited by F. B.M. de Waal and P.F. Ferrari), pp.81-98, Harvard University Press.
Visalberghi, E., and Fragaszy, D. (2013). The Etho-Cebus Project: Stone-tool use by wild capuchin monkeys. In *Tool Use in Animals: Cognition and Ecology (*eds C. Sanz, J. Call, C. Boesch), Cambridge University Press.

CNR-Barto

The question of what intrinsic motivation really means, and how it differs from extrinsic motivation, is controversial among psychologists as well as computer scientists and roboticists. Prof. Barto contributed a unique perspective to this question based on computational experiments from his group and colleagues that combined ideas from evolution and reinforcement learning. In these experiments, an automated search was performed to find out how a learning agent should be rewarded in order to have a 'lifetime' of high evolutionary fitness. It was found that in addition to rewarding the agent so as to encourage it to do what contributes in an obvious way to its lifetime fitness (analogous to rewards for eating, escaping, etc. in an animal) it was also good to reward it for doing things that were much less obviously related to lifetime success, but related nevertheless, such as exploratory activity and analogs of play. This evolutionary perspective led to the position that there are no hard and fast features distinguishing intrinsic and extrinsic reward. Instead, the directness of the relationship between rewarding behavior and evolutionary success varies along a continuum. What we call intrinsic rewards are rewards that are more distally related to reproductive success than more biologically-primal rewards. This perspective, which was extensively discussed by Singh et al. (2010), Barto, et al. (2011), and Barto (2013), was an important contribution to discussions of this issue by the IM-CLeVeR teams, especially ISTC-LOCEN.

A related focus of the UMass group was on the issue of using a measure of 'learning progress' to decide which skill to practice among a collection of partially learned skills. The UMass group made several important connections to existing areas of study. The extensive literature in psychology on 'metacognition', studies how humans monitor their own cognitive processes and use this information in deciding what to study next. Literature in this area explicitly considers assessing learning progress and how it influences decision-making. Another connection is to the theory of Gittins indices. The problem of switching practice among skills is closely related to the classic multi-armed bandit problem to which Gittins indices provide an optimal solution for switching among reward producing 'projects'. There are differences, but a thorough understanding of Gittins indices and related classical theory is useful in developing a better understanding of the possible role of intrinsic motivations in guiding behavior during skill learning. A related connection is to research in optimal foraging of behavioral ecology, where food patches for foraging animals are like specific skills, and eating is like practicing

Extensive collaboration with the ISTC-LOCEN group led to implementation of a sophisticated skill learning method using demonstrated on the i-Cub robot. This was carried out largely through the part-time residence in Rome of UMass graduate student Bruno Castro da Silva, as advised by Prof. Barto. This work is based on the theoretical framework called 'options' by machine learning researchers. Options provide a formal notion of what a skill is in the theory of reinforcement learning. Unlike actual skills, however, options are inflexible in each option accomplishes a single goal. In this work, the option framework was extended to 'parameterized options', which allows each option to be modulated by a parameter that can be set depending on specific demands of the task. This approach is described in Castro da Silva, et al. (2012). Subsequent to the publication of that paper, the method was implemented in an i-Cub throwing task in which the robot learned to accurately throw a ball to a variety of different spatial targets. These results will form a major part of Castro da Silva's UMass Ph.D. dissertation.

Barto, Mirolli, and Baldassarre collaborated on a paper entitled **"Novelty or Surprise?"** that was submitted for publication and is currently under review (Barto, et al. submitted). Novelty and surprise play significant roles in animal behavior and in attempts to understand the neural mechanisms underlying it. They are particularly relevant to intrinsic motivation, being among the primary factors that arouse interest, motivate exploratory or avoidance behavior, and drive learning. In many of these studies, novelty and surprise are not distinguished from one another: the words are used more-or-less interchangeably. However, while undeniably closely related, novelty and surprise are very different. The purpose of this paper is to highlight the differences between novelty and surprise, to discuss how they are related, and to explore the implications of this for understanding behavioral and neuroscience data. The authors argue that opportunities are likely being missed for improved understanding of behavior and its neural basis by failing to distinguish between novelty and surprise.

Selected Bibliography

Barto, A., Mirolli, M., and Baldassarre, G. (submitted) Novelty or Surprise?

Barto, A. (2013) Intrinsic motivation and reinforcement learning In G. Baldassarre and M. Mirolli, editors, *Intrinsically Motivated Learning in Natural and Artificial Systems*, pp. 17-47, Springer.

Castro da Silva, B., Konidaris, G.D., and Barto, A.G. (2012) Learning parameterized skills *Proceedings of the Twenty Ninth International Conference on Machine Learning*, pp. 1679-1686.

Barto, A.G., Singh, S., and Lewis, R.L. (2011) Intrinsically motivated machines *Proceedings of the Fifteenth Yale Workshop on Adaptive and Learning Systems*, Center for Systems Science, Department of Electrical Engineering, Yale University, pp. 123-129.

Vigorito, C.M. and Barto, A.G. (2010) Intrinsically Motivated Hierarchical Skill Learning in Structured Environments *IEEE Transactions on Autonomous Mental Development (IEEE TAMD)*. Volume 2, Issue 2, pp. 83-90.

Singh, S., Lewis, R.L., Barto, A.G., and Sorg, J. (2010) Instrinsically Motivated Reinforcement Learning: An Evolutionary Perspective *IEEE Transactions on Autonomous Mental Development (IEEE TAMD)*. Volume 2, Issue 2, pp. 70-82.

UCBM

May Curiosity be considered the driving force that shapes the process of acquisition of new skills and know-how? If not the driving force it could be definitely considered one of the key elements of this process. Curiosity drives subjects to spontaneously explore the environment providing them with an increasingly diverse set of opportunities for acquiring, practicing and refining new abilities. It is elicited and promoted by complex unexpected and surprising stimuli as suggested by Berlyne (Berlyne, 1966) and, probably, it might be used to address new knowledge in children and animals enabling a sort of *learning by doing* process completely different from classical operant conditioning paradigms. Recently works have investigated neural mechanism promoting this process. A key role seems to be played by the phasic release of dopamine caused by unexpected events. Redgrave and Gurney (Redgrave & Gurney, 2006) suggested this releasing is implicated in the process of discovering new actions. The architecture of this system seems to be common to different animals from humans to monkey to other less evolved forms of life. The goal of the IM-CLeVeR

project was to investigate such mechanisms that guide discovering and learning of new knowledge without any apparent external reward.
UCBM has a strong experience in robotics, neural engineering, and behavioural experiments with children. In particular, the Laboratory of Developmental Neuroscience (LDN) combines expertise in developmental neuroscience, neurobiology (specifically the neurobiology of sensorimotor systems), and psychology. It has pioneered a new approach to objective measurement of child behaviour in ecological scenarios (Campolo et al., 2012) (home, kindergarten etc.). The laboratory of Biomedical Robotics and Biomicrosystems (LBRB) has a solid background in human-centred design of robotic (Tagliamonte et al., 2011; 2012), mechatronic and smart systems (Taffoni et al., 2011; Di Pino et al., 2012), and in particular it matured relevant experience on the design and development of instrumented toys specifically devised for behaviour analysis of infants (Campolo et al., 2011).
UCBM has contributed to the project from a technological, experimental and theoretical point of view. In particular it has provided:

- a common technological tool to study different experimental populations: children, animal models and robotic platform. The proposed tool was designed in two release: one for children and robot (three replicas), one for monkeys (one replica);
- technical support in the data acquisition sessions both with children and monkeys;
- software tools developed in Matlab environment to assist UCBM-LDN for data analysis of children experiment;
- a net of contact that has allowed to involve 36 children in the experimental session. Twelve children were involved in a pilot study carried out during 2010-2011 to refine the experimental protocol and indexes. The other twenty-four subjects aged 3 and 4 year, were involved in the final experiment.
- Investigations and modelling of development of rhythmic manipulation skills.

To effectively study mechanisms of learning guided from curiosity it was necessary to develop a common technological tools enabling the study and comparison of different experimental populations (to verify if these mechanisms are common to different species). In particular, these processes have been studied in preschool children (three and four years of age) and Capuchin monkeys, a non human-monkey specie with a high level of manipulation thanks to its partially opposable thumb. These two populations represent good candidate to study mechanism of learning guided by curiosity and mediated by actions because they have a good basic motor repertoire. The approach used was phenomenological: in the presence of a sudden unexpected event related to the subject's action how he/she respond? Does he/she explore that particular action that caused the unexpected effect? Is he/she able to understand *what* has caused the effect? Is he/she able to understand spatial relationship between actions and effects (*where)?* How action-outcome contingency affect the behaviour? The neural mechanisms supposed to control these aspect of behaviour were modelled and used to control a child-like robot which was tested in the same experimental conditions and its behaviour compared with the ones measured for children and monkeys.
To enable this comparison, a common technological tool has been developed (Taffoni et al., 2013). Such a tool, hereafter called mechatronic board, is the result of a multidisciplinary design process which has involved bio-engineers (UCBM-LBRB), developmental neuroscientists (UCBM-LDN), primatologists (Consiglio Nazionale delle Ricerca, Unit of Cognitive Primatology & Primate Center, CNR-UCP), machine learning experts (Consiglio Nazionale delle Ricerca, Istituto di Scienza e Tecnologia della Cognizione, CNR-ISTC) and psychologists (University of Scheffield, USFD) to identify the main requirements and specifications of the technological platform as well as of the protocols. A hierarchical, three level architecture was adopted to

guarantee the reconfigurability and robustness of the platform. The platform was developed in two releases, one common to children and robot (Taffoni et al., 2012a), and one for monkeys (Taffoni et al., 2012b). It is composed by a planar base and a reward releasing unit. The planar base is provided of three slots where different smart-objects (i.e. objects instrumented with sensors to measure the users interaction) may be plugged-in while the reward releasing unit is mounted on the back area of the planar base and contains three reward boxes, where rewards may be placed by the experimenter. Various sensory feedbacks associated to the manipulation of mechatronic objects may be provided to subjects. The stimuli come both from the mechatronic objects (object stimuli) and from the reward releasing boxes (box stimuli), and may be completely reprogrammed.

A set of 12 subjects tested the protocol and the control condition. In the final version of the protocol developed by UCBM-LDN, the mechatronic board was equipped with the three simple pushbuttons which activate sounds, lights, and the opening of the three small doors in the reward releasing unit. Any button, if pressed just for a short time, activated only sounds and lights on the planar base, whereas keeping it pressed for a prolonged time (1 sec) activated the opening of one box, the lights and the speaker up to it. The experimental session involved three phases: Baseline, Learning and Test. These three steps were always presented in the same order for all subjects and the only phase that differed between the EXPerimental (EXP) and ConToRoL (CTRL) group was the Learning phase. During that, each subject of the CTRL group has been paired with one subject of the EXP group matched for age.

The baseline phase: this phase was administered as first one. The goal of this phase was to estimate the initial skills of each child and his/her interest in exploring the box. It lasted 5 minutes. During this phase, the only audio-visual stimuli came from the planar base and were contingent to a button press.

The learning phase: during this phase, children were allowed to explore the board and to discover how it worked "by chance". For the EXP subjects, the board was programmed to respond to each single button pushes with both visual and auditory stimuli and to open the reward boxes when a button was held pressed for more than one second. A Simple Push (SP) switched on the lights close to the pushed button and produced a sound of xylophone (three different tones are set for the three different buttons). The Extended Push (EP), instead, produced the opening of one box (rewarded action), which was always empty in this phase, the interior of the box lights up and the speaker near the box generates the sound of an animal (a different one for each button: a rooster, a frog and a cat). The CTRL children were yoked to the EXP ones: the mechatronic board was programmed to record the action of CTRL subjects and to deliver the outcomes of the actions performed by the paired EXP subjects. In this way CTRL subjects received the same number and the same kind of stimuli as the paired EXP subjects but independently from their actions. This condition make impossible for CTRL subject to learn any action-outcome relationship. This phase lasted 10 minutes.

Test phase: in this phase a sticker was used as reward. In this phase, the outcome depended on the action for both EXP and CTRL groups. The action-outcome relations were set as in the learning phase for the EXP group: the same settings were used for yoked CTRL subjects. Children were asked to retrieve the sticker placed in one of the three closed boxes in a clearly visible way. In order to avoid a bias effect due to the reward presentation order, three different sequences of the sticker position were defined. The sequences were randomly assigned and counterbalanced between the EXP subjects. The paired CTRL subjects received the same sequence order. The child was encouraged to retrieve the sticker without any other suggestion on what action was associated to the box opening. The test session consisted in 9 trials. Each trial began with the reward inside a box and finished within 2 minutes or before if the child retrieved the reward earlier. When the subject succeeded in opening the door and getting the reward, a new reward was placed inside another box. When the subject did not get the reward within two minutes, the same reward was placed inside another door. The testing session ended when the trials were concluded regardless of success in getting the rewards. In this phase, the outcome depended on the action for both EXP and CTRL group, so both groups might understand the action-outcome contingency. The goal of this phase was to verify if the child was able to exploit the skills

acquired in the learning phase in which participants should have understood that pressing a certain button a specific door was opened.
The above protocol was administered to 24 children by researched of UCBM-LDN with the technical support of UCBM-LBRB: 12 children aged three years (36.7 ± 0.8 months, mean ± Standard Deviation) and 12 children aged four years (48.5±0.8 months). For each age group, the recruited children have been randomly assigned to an EXP or a CTRL group.
The effect of action-outcome contingency in the exploration during the Learning phase seems to be age-related. There is no difference in terms of interaction with the board (expressed as number of pushes) between EXP and CTRL subjects in the younger group. On the other hand, four years old CTRL subjects loose their interest in the platform after five minutes and less frequently perform extended pushes than EXP subjects. The experience of a novel unexpected stimulus, like the box aperture, contingent with the action, the extended push, seems to promote the action learning independently from the age. Indeed, both three and four years old EXP children understand that an extended push causes the box opening. This can be argued by the negative correlation, only present in the EXP group, between the percentage of extended pushes in the Learning phase and the time to reward in the Test phase. No focussing behaviour has been clearly observed except for one experimental subject. A more abstract relationship, the spatial learning i.e. the understanding of the right association between the button and the box that it controls, seems to be learnt by EXP subjects during the Test phase. To assess this aspect, we have introduced the *spatial correctness* variable, defined as the ratio between the difference of right and wrong pushes, and the total number of pushes in each trial. We have grouped the 9 trials of the test phase in three triplets: each triplet presents two crossed associations and one direct association. We found that EXP subjects extend their learning during Test Phase improving the spatial correctness, which positively differs from 0 (random behavior) only from the second triplet. CTRL subjects do not learn spatial correctness during Learning Phase.
Additional efforts was dedicated in the modeling of fine manipulation. UCBM-LBRB has a strong expertise in modeling (Formica & Guglielmelli, 2012). In particular the It has been studied the role of thumb opposition during cyclic manipulation tasks through the interaction with different objects and a bio-inspired control architecture based on reinforcement learning. The control architecture has been implemented in simulated environment on two robotic hands with different thumb features, i.e. the iCub hand and the DLR/HIT Hand II, interacting with objects of different sizes and shapes (Ciancio et al., 2011; 2012; 2013).

Selected Bibliography

Berlyne, D. E. (1966). Curiosity and exploration. Science, 143, 25–3.

Redgrave, P. and K. Gurney (2006). The short-latency dopamine signal: a role in discovering novel actions? *Nature Reviews Neuroscience, 7*(12), 967–975.

Campolo, D., Taffoni, F., Formica, D., Iverson, J., Sparaci, L., Keller, F., & Guglielmelli, E. (2012). Embedding inertial-magnetic sensors in everyday objects: Assessing spatial cognition in children. Journal of Integrative Neuroscience, 11, 1, 103–116.

Tagliamonte, N., Scorcia, M., Formica, D., Campolo, D., and Guglielmelli, E. (2011). Effects of impedance reduction of a robot for wrist rehabilitation on human motor strategies in healthy subjects during pointing tasks. *Advanced Robotics, 25*, 537–562.

Tagliamonte, N., Formica, D., Scorcia, M., Campolo, D., and Guglielmelli, E. (2011). Force control of a robot for wrist rehabilitation: Towards coping with human intrinsic constraints. In *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, 2010, pp. 4384–4389.

Taffoni, F., Piervirgili, G., Formica, D., and E. Guglielmelli. An alignment procedure for ambulatory measurements of lower limb kinematic using magneto-inertial sensors. In *Engineering in Medicine and Biology Society,EMBC, 2011 Annual International Conference of the IEEE*, 2011, pp. 1197–1200.

Di Pino, G., Formica, D., Melgari, J., Taffoni, F., Salomone, G., di Biase, L., Caimo, E., Vernieri, F., and Guglielmelli, E. (2012). Neurophysiological bases of tremors and accelerometric parameters analysis.

In *Biomedical Robotics and Biomechatronics (BioRob), 2012 4th IEEE RAS EMBS International Conference on*, 2012, pp. 1820–1825.

Campolo, D., Taffoni, F., Formica, D., Schiavone, G., Keller, F., and Guglielmelli, E. (2011). Inertial-magnetic sensors for assessing spatial cognition in infants. *Biomedical Engineering, IEEE Transactions on,58*,5, 1499–1503.

Taffoni, F., Formica, D., Schiavone, G., Scrocia, M., Tomassetti, A., Polizzi di Sorrentino, E., Sabbatini, G., Truppa, V., Mannella, F., Fiore, V., Mirolli, M., Baldassarre, G., Visalberghi, E., Keller, F., and Guglielmelli, E. (2013). The "Mechatronic Board": a tool to study intrisic motivation in humans, monkeys, and humanoid robots, in Baldassarre, G.; Mirolli, M (Eds.) *Intrinsically Motivated Learning in Natural and Artificial Systems,* pp 411-432. Springer Berlin Heidelberg.

Taffoni, F., Formica, D., Zompanti, A., Mirolli, M., Balsassarre, G., Keller, F., and Guglielmelli, E. (2012a). A mechatronic platform for behavioral studies on infants. In *Biomedical Robotics and Biomechatronics (BioRob), 2012 4th IEEE RAS EMBS International Conference on*, 2012, pp. 1874–1878.

Taffoni, F., Vespignani, M., Formica, D., Cavallo, G., Polizzi di Sorrentino, E., Sabbatini, G., Truppa, V., Visalberghi, E., Keller, F., and Guglielminelli, E. (2012b). A mechatronic platform for behavioral analysis of nonhuman primates. *Journal of Integrative Neuroscience, 11*, 87-101.

Formica, D., and Guglielmelli, E. (2012). A basis fields approximation for modeling the passive elasticity of the wrist. *In Biomedical Robotics and Biomechatronics (BioRob), 2012 4th IEEE RAS EMBS International Conference on*, 2012, pp. 1247–1251.

Ciancio, A. L., Zollo, L., Baldassarre, G., Caligiore, D., and Guglielmelli, E. (2012). The role of thumb opposition in cyclic manipulation: A study with two different robotic hands. In *Biomedical Robotics and Biomechatronics (BioRob), 2012 4th IEEE RAS & EMBS International Conference on* (pp. 1092-1097). IEEE.

Ciancio, A. L., Zollo, L., Guglielmelli, E., Caligiore, D., and Baldassarre, G. (2011). Hierarchical reinforcement learning and central pattern generators for modeling the development of rhythmic manipulation skills. In *Development and Learning (ICDL), 2011 IEEE International Conference on* (Vol. 2, pp. 1-8). IEEE.

Ciancio, A. L., Zollo, L., Baldassarre, G., Caligiore, D., and Guglielmelli, E. (2013). The role of learning and kinematic features in dexterous manipulation: a comparative study with two robotic hands. In press on *International Journal of Advanced Robotic System*.

USFD

The University of Sheffield contribution to the IM-CLeVeR project was divided between experimental and computational modelling components, each receiving support for one post-doctoral research position. The principal achievements of each component of the Sheffield team will be documented separately.

Principal achievements of USFD experimental component

A fundamental problem for multifunctional autonomous agents is how to develop or acquire novel behaviour. Biological systems are intrinsically motivated to discover what aspects of their own behaviour are responsible for causing things to happen in the external world. The contribution of the experimental group in Sheffield was to provide insights into the mechanisms and processes used in biological systems in the discovery of agency and development of novel actions. This knowledge could then be used to inform the development of more adaptive and flexible artificial systems.

At the outset our first task was to develop an experimental paradigm with which to investigate novel action discovery. An important constraint on developing a suitable task was that it should be able to assess all the fundamental aspects of action acquisition – WHERE and action has to be performed, WHAT manipulation has to be made, WHEN it has to be made, and HOW it has to be made. With these constraints in mind we have developed various versions of a joy-stick task (Stafford et al., 2012). The task consists of free movement with a manipulandum, during which the full range of possible movements can be explored by the participants and recorded. A sub-set of these movements, the 'target', is set to trigger a reinforcing signal. The subject has to discover what movements of the manipulandum evoke the reinforcement signal. This task is sufficiently versatile that it can be used to study the different aspects of action independently. It can be implemented on different physical platforms (joy-stick, touch screen, computer mouse/tracker-ball, and gaze direction) and performed by a wide range of subjects (rodents, monkeys, humans and robots). Depending on the criteria chosen for reinforcement its difficulty can be continuously varied, and by changing the reinforcement criteria, it can be used in experiments requiring repeated measures. Our joy-stick task has been used extensively in our work for the ImClever project and has been transferred directly to many other laboratories, including our animal laboratory in Sheffield also in New Zealand (rats), Germany (human infants), Spain and Japan (non-human primates), and Ireland (human patients). Finally this task produces a rich set of behavioural measures offering new and valuable insights into the process of learning novel actions. It is interesting to note that a modified version of the task was recently developed independently by a top American group (Chukoskie et al., 2013).

We have used the joy-stick task with humans to explore several important aspects of novel action discovery. First, we were able to show that subjects who explored more during the early period of learning ultimately achieved higher levels of performance (Stafford et al., 2012). This principle can be incorporated into learning algorithms used to acquire novel actions in artificial systems. Secondly, we used the task to investigate the biological basis of visually reinforcing stimuli (Thirkettle et al., 2013b). It is widely assumed that phasic activity of the neurotransmitter dopamine (DA), which is evoked by unpredicted salient visual events, provides the critical reinforcement that links recently performed movements with their outcome. However, at the beginning of the ImClever grant the latency of the visually evoked DA response excluded any significant contribution from the sophisticated visual processing provided by cerebral cortex. However, by using stimuli that were unavailable to the less sophisticated sub-cortical visual systems we showed that stimuli processed exclusively by cerebral cortex do in fact have access to the brain's basic reinforcement mechanisms. Although it was clear that stimuli that were available to both cortical and sub-cortical visual systems (luminance changes) were more effective reinforcers than stimuli available to cerebral cortex only (colour changes). This is important because it shows that a wide range of visual events can be used to reinforce intrinsically and extrinsically motivated action acquisition, not just those that can be processed sub-cortically. Third, we explored the efficacy of visual compared with auditory reinforcers and various time-delays between the 'correct movement' and the consequent reinforcing signal (Walton et al., 2013). Auditory reinforcement was found to be superior to visual and that reinforcer delays of only 75-150 milliseconds significantly impaired the learning of novel actions. This study shows that in tasks that are continuously dynamic the credit assignment problem (which aspect of recent behaviour causes the reinforcing outcome) is exacerbated and that even short delays are detrimental. This suggests that when artificial agents are required to learning dynamic tasks it is essential to minimise reinforcement delay. Fourthly, we performed an experiment to investigate the relative ease of learning WHAT to do, as opposed to WHERE to do it (Thirkettle et al., 2013a). In this study we asked subjects to use the joy-stick to search for a hidden target area from fixed and variable start positions. We showed it was much easier to learn the location of the hidden area from a fixed start point. This shows that for biological systems the WHAT (trajectory) aspect of a novel action is easier to learn than the WHERE (movement endpoint) aspect.

An additional study was proposed in the original application that involved novel action acquisition in patients with Parkinson's disease. Following the preparation and publication of what has turned out to be an influential review of the action-outcome learning in Parkinson patients (Redgrave et al., 2010) the original study was discovered to be inappropriate and therefore shelved. It was replaced by a collaborative study with a Japanese group using a gaze-direction version of the joy-stick task in monkeys. This study, which is currently being prepared for publication, confirmed our human study (see above) by showing that visual reinforcement that engages both cortical and subcortical sensory processing is more effective than, in this case, sub-cortical visual processing alone.

Selected Bibliography

Chukoskie L, Snider J, Mozer MC, Krauzlis RJ, Sejnowski TJ (2013) Learning where to look for a hidden target. Proc Natl Acad Sci U S A.

Redgrave P, Rodriguez M, Smith Y, Rodriguez-Oroz MC, Lehericy S, Bergman H, Agid Y, DeLong MR, Obeso JA (2010) Goal-directed and habitual control in the basal ganglia: implications for Parkinson's disease. Nat Rev Neurosci 11:760-772.

Stafford T, Thirkettle M, Walton T, Vautrelle N, Hetherington L, Port M, Gurney K, Redgrave P (2012) A novel task for the investigation of action acquisition. Plos One 7:e37749.

Thirkettle M, Walton T, Redgrave P, Gurney KN, Stafford T (2013a) No learning where to go without first knowing where you're coming from: action discovery is trajectory, not endpoint based. Frontiers in Cognitive Science (in press).

Thirkettle M, Walton T, Shah A, Gurney K, Redgrave P, Stafford T (2013b) The path to learning: action acquisition is impaired when visual reinforcement signals must first access cortex. Behav Brain Res 243:267-272.

Walton T, Thirkettle M, Redgrave P, Gurney KN, Stafford T (2013) The Discovery of Novel Actions Is Affected by Very Brief Reinforcement Delays and Reinforcement Modality. J Mot Behav.

Principal achievements of USFD modeling component

- A computational model of the experimental paradigm described in the experimental section above (the joystick task developed at USFD) which showed how the basal ganglia, together with its associated brain strictures, could learn a ‘map’ of the workspace in the task in which a specific spatial location (or locations) were reinforced. This was therefore a model which learned *purely* the WHERE component of an action. Our model of the joystick task and its predictions has sharpened our conceptual thinking about this task and, quite generally, about the interaction between the intrinsically motivated learning of different action components in realistic situations. While the components (WHERE, WHAT etc) may have separate encodings in the brains, they will, in general be present in most behavioral situations. We have also been led to consider the nature of exploration during intrinsically motivated learning and the interaction between cognitive ‘strategic’ exploration and random exploration. Our model of the joystick task only uses the latter, and testing the model experimentally (already underway) will require suppression of strategic planning in humans.

- We garnered new insights into the reasons why the neurotransmitter dopamine (a key factor in learning new actions) behaves the way it does. That is, the amplitude of the dopamine signal gradually decreases with increased task performance. Failure to do so with simple learning rules of the kind we expect to occur in the brain leads, eventually, to an ‘unlearning’ of the task.

- A model of learning new actions in an autonomous (virtual) robot which could account for experimental results in animals, This model prompted a new hypothesis for action learning (the associated objects and situations become intrinsically ‘interesting’ and ‘salient’ for a while) which will be investigated in future work. Several ideas in this work were used in the main project demonstrator (Clever-B). This model has helped improve our conceptualization of learning new actions which may require the ‘gluing together’ of sensory context and, either an existing ‘atomic’ action, or a combination of such action components for a new skill.

- A hierarchical model of multiple basal ganglia-loops for sequential action generation. This model combines action transitions in prefrontal cortex with their execution in motor cortex and can account for a raft of motor errors observed in humans.

Selected Bibliography

Bolado-Gomez, R., & Gurney, K. N. (2013). A biologically plausible embodied model of action discovery. *Frontiers in Neurorobotics*, *7*(4). doi:10.3389/fnbot.2013.00004

Gurney, K. N., Lepora, N., Shah, A., Koene, A., & Redgrave, P. (2013). Action discovery and intrinsic motivation: a biologically constrained formalisation. In *Intrinsically Motivated Learning in Natural and Artificial Systems.* Springer.

Lepora, N. F., & Gurney, K. N. (2012). The basal ganglia optimize decision making over general perceptual hypotheses. *Neural Computation*, *24*(11), 2924–2945.

Humphries, M. D., Khamassi, M., & Gurney, K. (2012). Dopaminergic control of the exploration-exploitation trade-off via the basal ganglia. *Frontiers in Decision Neuroscience*, *6:*, 9. doi:10.3389/fnins.2012.00009

Baldassarre, G., Mannella, F., Fiore, V. G., Redgrave, P., Gurney, K., & Mirolli, M. (2012). Intrinsically motivated action-outcome learning and goal-based action recall: a system-level bio-constrained computational model. *Neural Networks, 41*, 168-187.

Shah, A., & Gurney, K. (2011). Dopamine-mediated action discovery promotes optimal behavior “for free.” *BMC Neuroscience*, *12*(Suppl 1), P138.

Shah, A., Thirkettle, M., Tidman, J., & Gurney, K. N. (2013). Emergent variation and repetition trade-off in a biologically-plausible model of intrinsically motivated action discovery. *Frontiers in Cognitive Science*. (submitted)

Shah, A. S., & Gurney, K. N. (2013). Dopamine-mediated Action Discovery Promotes Optimal Behavior “for Free.” *PLoS Comp Bio* , (MS prepared, to be submitted).

FIAS

Major contributions of the FIAS team were achieved in a number of areas:

A curious vision system for autonomous object learning

The ability of biological organisms to learn autonomously is a challenge for future robots assisting humans in their homes, work places, or other changing environments. Autonomous learning implies that the system can choose what activities it should engage in and hence what it can learn about. Central to these activities are learning and identifying the objects in a scene. There are many object learning systems in the literature that are trained with supervised classification techniques. In contrast, we introduced an autonomously learning robot vision system implemented on the iCub robot, that explores a scene and learns visual object representations without human assistance. The system integrates a number of visual competencies including attention, stereoscopic vision, segmentation, tracking, model learning, and recognition. The system in particular is inspired by the concept of intrinsic motivations (Baldassarre, 2011). Similar to an infant, who is intrinsically motivated to seek out new information, our system is endowed with an attention mechanism designed to search for new things to learn about. New object models are created if objects are judged to be new, and existing object models are updated if objects are judged to be familiar. The system continues to learn about an object as long as the model of the objects can be further improved by acquiring new features from the object. The attention is diverted when the learning progress drops to a small value and the process starts over. This work has been pursued in collaboration with the IDSIA and CNR labs and has led to a number of publications (Chandrashekhariah et al., 2013; Leitner et al., 2012, 2013). A Master student from UCBM (Gabriele Spina) was also actively involved.

Link to the video clip depicting system performance : http://youtu.be/qdYzlnMtjIw

Autonomous Learning of Binocular Vision

The efficient coding hypothesis posits that sensory systems of animals strive to encode sensory signals efficiently by taking into account the redundancies in them. This principle has been very successful in explaining response properties of visual sensory neurons as adaptations to the statistics of natural images. In this work, which is in collaboration with Bert Shi's lab at the Hong Kong University of Science and Technology, we extend the efficient coding hypothesis to active perception through a form of intrinsically motivated learning. The method makes use of a form of intrinsic motivation to learn efficient sensory representations in the perception-action cycle. A sparse coding model encodes sensory information using binocular basis functions, while a reinforcement learner generates the eye movements to aid in the encoding of the sensor signals. Importantly, perception and behavior develop in parallel, by minimizing the same cost function: the reconstruction error of the stimulus by the generative model. The system autonomously discovers that it is useful to make proper vergence eye movements, because this allows to encode the sensory signals more efficiently. A multi-scale approach, where basis functions are learned at different resolutions, is used to cope with multiple disparity ranges and has been demonstrated on the iCub robot (Lonini et al., 2013). The robustness of the approach was also tested thoroughly (Lonini et al., submitted). In addition, we extended the framework to motion perception and smooth pursuit (Zhang et al., submitted). Most recently, we also established a connection to imitation learning (Triesch, submitted).

Link to the video clip of the binocular vision system running on iCub: http://youtu.be/hGsvAuIFySc

Intrinsic motivations in a gaze-Contingent paradigm

As an infant tries to make sense of the vast array of signals from its sense organs and wins control over her body and physical environment, a fundamental challenge is to learn which sensory events are the consequence of her own motor actions and which ones are not: in other words, to discover agency. It has been difficult to shed light on this ability in infants because of their limited motor repertoires. We employed a novel gaze-contingent paradigm to overcome this limitation. We demonstrated that six to eight-month-old infants very quickly learn to selectively take actions to manipulate their environment and learn to anticipate the outcomes of their actions. We found that infants quickly learn to perform eye movements to trigger the appearance of new stimuli and that they anticipate the consequences of their actions in as few as 3 trials (Wang et al., 2012). Our findings showed that infants can rapidly discover new ways of controlling their environment. To better understand the underlying neural mechanisms, we have collaborated with the CNR team to develop a computational model of these findings (Marraffa et al., 2012).

Development of coordinated eye and head movements in primates

Shifts of attention are usually reflected by gaze shifts, which are characterized by coordinated eye and head movements in primates. Although much effort has been put forth to reveal the neural foundation underlying the gaze shift control system or to find the optimality principle behind its dynamics, it remains unclear how such a control scheme develops in primates. The goal of this research was to find an answer to the following fundamental questions: First, why coordinated eye and head movements have the stereotypical characteristics as observed in the experiments. Second, how does the control circuitry in the brain learn such a control task. We developed a computational model using a simple cost function and a plausible neural learning model (Saeb et al., 2012). The simulation results showed that our model is able to reproduce many of the gaze shift characteristics observed in experiments, both in head-restrained (HR) and head-free (HF) conditions. These characteristics include the so-called saccadic main sequence and the bell-shaped velocity profiles in HR, and the relative contribution of eye and head to the gaze shift as well as its dependence on initial eye position in HF conditions. Furthermore, our model reproduces the early fixation phase of the eyes in HR and the so-called vestibulo-ocular reflex (VOR) phase in HF conditions. All of these results are gradually achieved through incremental learning, and the model can be generalized to other ballistic motor control tasks beyond gaze shifts by finding appropriate motor plant models and cost functions.

Directed exploration and motivations

When an agent acts within a partially observable environment, it is crucial that it can direct its attention to environmental features that will give the maximum information relevant to the task (Taylor and Rogers, 2002; Seth, 2000; van de P. Laar et al., 1997). This is generally a hard decision theoretic problem (Ross et al., 2008; DeGroot, 1970). One special version of this problem is visual attention (Corchs and Deco, 2000). Related to this is the question of whether it is possible to learn heuristic exploration strategies for visual attention tasks and whether such strategies can be generalized to more general exploration problems. The final open issue is how to most effectively take the interaction between tasks and attention into account. A goal was to develop formal models, which have low computational and sample complexity. Our work progressed along several lines. We used context models (Dimitrakakis, 2011, 2010c) for the predictive modeling of features. In addition, we addressed the question of whether formal settings exist under which curiosity heuristics arise naturally as near-optimal solutions (Rothkopf and Dimitrakakis, 2011). We also had results on hierarchical modelling (Dimitrakakis, 2011, 2010c,b), where it is shown that simple, closed-form, hierarchical distributed inference can be used for prediction and action in a number of standard partially observable problems. In addition, we obtained results on approximate planning under uncertainty (Dimitrakakis and Lagoudakis, 2008; Dimitrakakis, 2010a), where we derive nearly optimal planning algorithms that use a relatively small amount of computation. Finally, we have some experimental and theoretical results linking intrinsic motivations and attention to decision theory through an interesting class of abstract problems and inverse reinforcement learning (Rothkopf and Dimitrakakis, 2011; Dimitrakakis et al., 2011; Dimitrakakis and Rothkopf, 2012).

Reinforcement learning for hierarchical decompositions in weakly-coupled problems

The goal of this work was to decompose complex reinforcement learning tasks in a hierarchical fashion. We have proposed a hierarchical reinforcement learning solution for the case in which the dynamics of different tasks are not independent but weakly coupled and show how to assign credit to the different modules, which solve the task jointly. We showed that a solution for such a task, where transition dynamics of one module affect that of another, can be obtained with a hierarchical RL approach. Individual model-free modules learn task solutions for individual dimensions. An additional model-free module furthermore learns to coordinate the independent lower-level modules by compensating for the weak coupling in their dynamics. We compared different structural credit assignment rules and training schedules and showed the significant differences in learning times and asymptotic performance through simulations (Toutounji et al., 2011).

Overall, the FIAS team has contributed to a number of advances regarding intrinsically motivated learning in the visual and other domains and developed novel cumulative and hierarchical learning schemes.

Selected Bibliography

Chandrashekhariah, P., Spina, G., & Triesch, J. (2013) Let it Learn: A curious vision system for autonomous object learning. *VISAPP - International conference on computer vision theory and applications, 21-24* February, 2013, Barcelona, Spain.

DeGroot, Morris H (1970). *Optimal Statistical Decisions*. John Wiley & Sons.

Dimitrakakis, C. (2010). Complexity of stochastic branch and bound for belief tree search in Bayesian reinforcement learning. *In 2nd international conference on agents and artificial intelligence (ICAART 2010), pages 259–264*, Valencia, Spain, 2010a. ISNTICC, Springer.

Dimitrakakis, C. (2010b) Bayesian variable order Markov models. *In Yee Whye Teh and Mike Titterington, editors, Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 9 of JMLR : W&CP, pages 161–168, Chia Laguna Resort, Sardinia, Italy.

Dimitrakakis, C. (2010c). Context model inference for large or partially observable MDPs. *In ICML workshop on reinforcement learning and search in very large spaces*.

Dimitrakakis, C. (2011). Context models on sequences of covers. *Technical Report arXiv:1005.2263v2, FIAS*.

Dimitrakakis, C., & Lagoudakis, M. G. (2008). Rollout sampling approximate policy iteration. *Machine Learning, 72(3):157–171*, doi: 10.1007/s10994-008-5069-3.

Dimitrakakis, C., & Rothkopf, C. A. (2012). Bayesian multitask inverse reinforcement learning. *Recent Advances in Reinforcement Learning, in Lecture Notes in Computer Science Volume 7188*, 273-284.

Leitner, J., Chandrashekhariah, P., Harding, S., Frank, M., Spina, G., Förster, A., Triesch, J., Schmidhuber, J. (2012). Autonomous Learning Of Robust Visual Object Detection And Identification On A Humanoid. *IEEE Conference on Development and Learning (ICDL) and Epigenetic Robotics (EpiRob)*. San Diego, USA. November 2012.

Leitner, J., Harding, S., Chandrashekhariah, P., Frank, M., Förster, A., Triesch, J., Schmidhuber, J. (2013). Learning Visual Object Detection and Localisation Using icVision. *Biologically Inspired Cognitive Architectures, Vol. tba*.

Lonini, L., Zhao, Y., Chandrashekhariah, P., Shi, B., & Triesch, J. (accepted). Autonomous learning of active multi-scale binocular vision. *To appear in proceeding of ICDL-EPIROB 2013*.

Lonini, L., Dimitrakakis, C., Rothkopf, C. A., Triesch, J. (accepted). Generalization and Interference in Human Motor Control. to appear in *Computational and Robotic Models of the Hierarchical Organization of Behavior*, Springer – Verlag, Berlin.

Lonini, L., Forestier, S., Teuliére, C., Zhao, Y., Shi, B. E., Triesch, J. (submitted to Frontiers in Neurorobotics, special Issue on: "intrinsic motivations in Humans, animals and robots). Robust Active Binocular Vision through Intrinsically Motivated Learning.

Marraffa, R., Sperati, V., Caligiore, D., Triesch, J., Baldassarre, G. (2012). A bio-inspired attention model of anticipation in gaze-contingency experiments with infants. *IEEE International Conference onDevelopment and Learning and Epigenetic Robotics (ICDL), 2012*.

Rothkopf, C., & Dimitrakakis, C. (2011). Preference elicitation and inverse reinforcement learning. *In Dimitrios Gunopulos, Thomas Hofmann, Donato Malerba, Michalis Vazirgiannis (Eds.) Machine Learning and Knowledge Discovery in Databases Part III: European Conference, Ecml Pkdd 2010*, Athens, Greece, September 5-9, 2011, Proceedings, Part III. Springer-Verlag.

Saeb, S., Weber, C., Triesch, J. (2011). Learning the optimal control of coordinated eye and head movements. *PLoS Computational Biology, 7(11), 2011, e1002253*.

Saeb, S., Weber, C., Triesch, J. (2011). Development of coordinated eye and head movements during gaze shifts. *Proc. Vision Sciences Society Annual Meeting (VSS'11)*, Naples, Florida, USA, 2011.

Toutounji, H., Rothkopf, C. A., & Triesch, J. (2011). Scalable reinforcement learning through hierarchical decompositions for weakly-coupled problems. *Proceedings of the IEEE International Conference on Development and Learning (ICDL-EpiRob-2011)* Frankfurt, 24 August, 2011.

Triesch, J. (submitted to Frontiers in Neurorobotics, special Issue on: "intrinsic motivations in Humans, animals and robots). Imitation Learning Based on an Intrinsic Motivation Mechanism for Efficient Coding.

Wang, Q., Bolhuis, J., Rothkopf, C.A., Kolling, T., Knopf, M., Triesch, J. (2012). Infants in Control: Rapid Anticipation of Action Outcomes in a Gaze-Contingent Paradigm. *PLoS ONE 7(2), e30884*. doi:10.1371/journal.pone.0030884.

Zhao, Y., Rothkopf, C. A., Triesch, J., & Shi, B. (2012). A unified model of the joint development of disparity selectivity and vergence control. *Proceedings of the IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob-2012)*, 7-9 Nov 2012.

Pengsheng, Z., Dimitrakakis, C., & Triesch, J. (2013). Network self-organization explains the statistics and dynamics of synaptic connection strengths in cortex. *PLoS computational biology, 9(1):e1002848*.

UU

In the IMCLeVeR project, University of Ulster consistently strived towards the development of cognitive robotics based on biological signal processing, and in particular intrinsically motivated decision making. In summary, during the first year significant effort was put into developing core novelty detection methods as precursors for intrinsic motivation. Following on from this, in year two, visual perception was used to extend novelty detection methods, in order to identify and learn objects and their associated actions. In year three the main goals were to explore effects of robot's interaction with objects for action learning, and thus allowing the robot to learn and acquire skills. The efforts were put in the final year to focus on extending skills learning and adaptation; specifically bridging the gap between neuroscience and computing through practical robotic demonstrations of intrinsic motivation models.

The main focus of UU was the development of core **novelty detection** methods, where novelty detectors were based on biological learning and habituation; also hierarchical representations of image data were investigated:

- A new approach for segmentation of objects using live streams from robot cameras in natural scenes in 3D was developed. An online method of detecting features and recognizing objects for the novelty detector was developed. Techniques were implemented to enable a physical robot to carry out actions on perceived objects and identify the outcome of these actions so that basic affordances of the objects can be associated with particular events.
- In addition cumulative learning systems, which were intrinsically motivated by novelty detection modules based on the biological non-associative learning form of habituation, were investigated. Existing novelty detection approaches were extended by integrating current state of the art visual object recognition algorithms. Specifically UU looked at characteristics needed for a novelty detector to be effective in cumulative robot learning (Gatsoulis et al 2010). UU developed a new novelty detector learner, validating the models in experiments with physical robots in real world environments.
- UU then implemented a Bayesian based method for hierarchical representation of image data and demonstrated that the approach is an effective method for information storage for a hierarchy of images. Also an evolutionary approach was developed for a robot to create a new skill and capability based on an appropriate combination and sequencing of lower lever skills.

UU then focused on extending core novelty detection methods using **visual perception processing** techniques**;** identification and learning of objects and associated actions based on visual perception. Vision is the most obvious, biological and potentially quickest way of learning further characteristics of an object:

- UU demonstrated a robot system that is capable of continuously and autonomously exploring and learning its perceptual state space, by identifying the most novel objects within its search space (as identified by visual processing and habituation) and learning their visual description (Gatsoulis et al., 2011a). Hence we demonstrated the limitations of the traditional bag-of-words and previous approaches in novelty detection when used in a cumulative robot learning scenario.
- UU developed a novel learning architecture, an expandable bag-of-words, for effective cumulative robot learning of visual perceptions (Gatsoulis et al., 2011b).

UU's next research goals were to explore effects of robot's interaction with objects for **action learning**; to focus on how a robot performs action learning of skills; to complement the visual perception novelty detection methodology by extension to the tactile domain:

- Action learning based on previously published expandable bag-of-word methods (Gatsoulis et al, 2011b) were extended with biologically inspired novelty detection for effective exploration and continuous robot learning (Gatsoulis et al., 2012a).
- A fuzzy neural network was implemented, which was used to learn and optimise basic affordances through interactions with an object (Gatsoulis et al., 2012b).
- A related issue was that of endowing a robot with the capability to autonomously combine, adapt and create actions to solve a task. UU reported work related to automatic composition and parameterization of skills; learning and adaptation of skills based on a novel evolutionary algorithm (Riano and McGinnity, 2012a; 2012b; 2012c).
- UU investigated combining tactile and visual information, noting that the tactile approach provides additional information related to weight, compressibility, surface texture, temperature and complicated contours in the exterior shape.
- UU also explored identification by examining objects using a Shadow Robotics hand with its in-built force strain gauges (Ratnasingnam and McGinnity, 2011a; 2011b; 2011c).
- UU implemented a robotic hand detection algorithm based on a well-known work to assist AU in some of their tasks as an important collaboration and integration across the project partners.

UU's research in the final year was dedicated to extend work on skills learning to enable **skills adaptation**; to progress from static to dynamic novelty detection and specifically to bridge the gap between the neuroscience and computing aspects of the project team through practical robotic demonstrations of **intrinsic motivation** models.

- In line with the overall ethos of the IM-CLeVeR project UU developed collaborations with specific neuroscience partners on the project, as exemplified through UU's work on integrating the CNR model of intrinsic motivation (Baldassarre et al., 2012) and extending it to action learning frameworks.
- UU focused on an experimental implementation and two extensions of the CNR intrinsic motivation model in the robot lab using a PR2 robot. UU devised an experimental setup for learning action consequences where the robot interacted with balls on a table with holes and had a limited selection of random actions. Vision modules enabled the robot to locate an object and use intrinsic motivations to learn to focus and track it. The robot was able to learn representations for various objects in its environment based on intrinsic motivations.
- A first extension to CNRs model involved a probabilistic biased selection approach (PBS) based on former acquired knowledge. This PBS approach is compatible with CNR's models and principles. Results from the new integrated approach showed consistent improved behavior and clear benefits. Inspiration was also taken from IDSIA's prediction theory in a further extension to the model where a predictor learnt and improved over time. The system was allowed to learn actions that had no result (as well as those which did have a result), hence improving the prediction abilities of the system.

To learn in a continuously changing environment, a novelty detector needs to be trained in an on-line manner in order to learn robot perceptions. UU's novelty detection methods used Radial Basis Function based recurrent neural networks to predict future sensory-motor readings (Ozbilge, 2011). It also learnt to estimate the current robot state in order to select associated state dependent local novelty threshold. The network approach is capable to be trained on-line but it is possible to change the network structure while receiving novel information from the environment.

In a biologically inspired novelty detection mechanism, an Evolving Connectionist Systems (ECoS) based network is used as a starting point. This type of network grows dynamically in the hidden layer by inserting a new node whenever a new input or predicted output does not match with an already existing network structure. Further, the network consists of a short-term memory, which allows the network to deal with temporal dependencies between consecutive input data. For detecting novelties in an environment, the ECoS network was modified such that it learns local novelty thresholds for each distinct input space simultaneously during the training. Thus, while the robot patrols, the predictions of the network can be evaluated differently in each local area in which the robot is currently travelling, with associated local thresholds. Specifically the network is capable of learning state-dependent local novelty thresholds simultaneously on-line during a training phase (Ozbilge, 2013)

Successful tactile exploration of an object for identification can assist visual processing and novelty identification. In terms of tactile exploration, UU explored an approach, which exploits simple force sensors mounted on a Shadow Robotics Hand. The aim was to construct a surface exploration algorithm for a Shadow Hand mounted on a robotic (Schunk manipulator) arm utilising simple and cheap tactile sensors only. Unlike many of the approaches presented in the literature, where often expensive tactile array sensors and vision systems are utilised (or required) to identify an object, this work used a force sensor with a single point contact reading the magnitude of the force on the tip of each finger. This algorithm was used to explore an object by exploring its outer surface by the use of tactile input only, in three dimensions. The approach used the positions of contact in 3D space to characterise the shape, meaning it can learn shapes of different sizes and curvatures. The characteristics of the shape are learnt by a neural network, so that the system can recognise if it encounters the same shape in future explorations. The method was tested for its capability to recognise differences in shapes due to tactile exploration of entire 3D shapes and to identify a shape from one point of view only by taking a 'slice' of the object, i.e. 2D, either by vision or tactile input (Kerr 2012).

More recent work utilises newly purchased advanced biomimetic tactile sensors BioTac™ produced by Syntouch®, for the classification of surface textures of objects and materials. The NN was also tested for robustness across different levels of data input. An action carried out by the robot within two different contexts may have different outcomes. The BioTac tactile sensors is shaped like the human fingertip and is liquid filled, giving it similar compliance to the human fingertip. Like human fingers, it is capable of detecting the full range of cutaneous sensory information: forces, micro vibrations and temperature. Using the fingertip, the robot contacts the surface on which the experiment is being carried out and slides along it to collect pressure and vibration data. Current work at UU focuses on extracting features from this filtered data to classify between surfaces, extending previous work at UU by Ratnasingham et al. (2011b).

Based on the outcome of a surface being rough or smooth, it is possible to integrate texture information into the environmental context of the CNR intrinsic motivation model, the robot could learn the effect of the surface within the context of the experiment, which will allow it to determine that to push the ball the same distance for a rougher surface, more force is needed than it would take on a smoother surface.

Selected Bibliography

Baldassarre, G., Mannella, F., Fiore, V. G., Redgrave, P., Gurney, K., & Mirolli, M. (2012). Intrinsically motivated action-outcome learning and goal-based action recall: a system-level bio-constrained computational model. *Neural Networks, 41*, 168-187.

Gatsoulis, Y., Burbridge, C. and McGinnity, T.M. (2011a) Online Unsupervised Cumulative Learning For Life-Long Robot Operation. In: 2011 IEEE International Conference on Robotics and Biomimetics (ROBIO), Phuket, Thailand. 2012 IEEE International Conference on, pp. 2486-2490

Gatsoulis, Y., Burbridge, C., McGinnity, T.M., (2012a) Biologically inspired intrinsically motivated learning for service robots based on novelty detection and habituation, Robotics and Biomimetics (ROBIO), 2012 IEEE International Conference on, pp.464,469, 11-14 Dec. 2012.

Gatsoulis, Y., Kerr, E., Condell, J.V., Siddique, N.H., McGinnity, T.M. (2010) Novelty detection for cumulative learning. Proc. of Towards Autonomous Robotic Systems 2010 (TAROS'10), p.62-67, 2010.

Gatsoulis, Y., Kerr, E., Condell, J.V., Siddique, N.H., McGinnity, T.M. (2011b) A visual novelty detection filter based on bag-of-words and biologically-inspired networks. Proc. of the AISB'11 Convetion - Symposium on Architectures for Active Vision, p.29-32, 2011.

Gatsoulis, Y., Siradjuddin, I., McGinnity, T.M. (2012b) Primitive action learning using fuzzy neural networks, Robotics and Biomimetics (ROBIO), 2012 IEEE International Conference on, pp.1513,1517, 11-14 Dec. 2012.

Kerr, E. (2012) Object Exploration and Identification System using Simlpe Force Sensors. MSc Thesis Report. University of Ulster, ISRC. 2012.

Ozbilge E., McGinnity, T.M., Condell, J., (2011), Novelty Detection for Autonomous Inspection Robot in Dynamic Environment, in Artificial Intelligence and Cognitive Science, 2011, AICS 2011. 22nd Irish Conference on AICS, Aug. 2011, pp. 224–233.

Ozbilge, E., Expectation-Based Novelty Detection For Mobile Robots(2013), Ph.D. dissertation, Faculty of Computing and Engineering, University of Ulster, Londonderry, July 2013.

Ratnasingam, S., McGinnity, T.M., (2011a), Object recognition based on tactile form perception, 2011 IEEE SSCI Workshop on Robotic Intelligence in Informationally Structured Space (RIISS 2011), 26-31; ISBN-13: 978-1-4244-9885-7; DOI: 10.1109/RIISS.2011.5945777.

Ratnasingam, S., McGinnity, T.M. (2011b), A spiking neural network for tactile form based object recognition, 2011 International Joint Conference on Neural Networks (IJCNN 2011 - San Jose), p 880-885.

Ratnasingam, S., McGinnity, T.M. (2011c), A comparison of encoding schemes for haptic object recognition using a biologically plausible spiking neural network 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2011), p3446- 3453.

Riano L and McGinnity T.M., (2012a). Automatically Composing and Parameterizing Skills by evolving Finite State Automatae, Robotics and Autonomous Systems, 60 (4). pp. 639-650

Riano L and McGinnity T.M., (2102b) A Robot that Autonomously Improves Skills by Evolving Computational Graphs, Accepted for W CCI 2012, Brisbane Australia.

Riano L and McGinnity T.M., (2012c) Autonomous Skills Creation and Integration in Robotics, AAAI Symposium USA 2012.

AU

Infant & Staged Development

During the IM-CLeVeR project we furthered our theories surrounding behavioural and cognitive development in robotics and have established a framework for implementing development on robotic platforms (Law et al., 2011). Early infancy provides a rich source of inspiration for developing algorithmic models that may allow robots to learn autonomously. Infants develop through a series of behavioural stages, showing the cycle of learning and generalisation of competencies that will support the infant during its lifetime. Behaviours rapidly emerge, consolidate, are superseded, or fused together creating new and improved competencies, during a period of intense activity and change. Although stages and their timings vary between individuals, it is widely recognised that learning progresses through a sequence influenced both by internal and external factors.

Our approach is firmly grounded in the study of infant development, and describes how data from infant psychology can be mapped onto a robotic system. The robot develops along an infant-like trajectory, learning similar behaviours and skills at appropriate stages. We achieve this by employing a series of constraints on the robots sensors, motor systems and cognitive layers. These constraints are lifted, one by one, as capabilities are sufficiently learned. We have explored how the use of constraints relate to staged behaviour and competence learning.

The importance of constraints has been shown in shaping robotic learning and reducing complexity, enabling fast, online learning (Lee et al., 2012). Constraints identified in the infant literature provide grounds for robotic constraints that enable effective shaping of learning. We have shown how different forms of constraint impact on learning, and direct the robot to learn useful behaviours based on its existing ability and environment.

To investigate the transition between developmental stages in response to triggered and emergent constraint release, we examined two types of constraint (Law et al., 2013): Type A constraints are created by immaturity in the growing neural, physiological and bodily structures, and are removed from the relevant systems when sufficient physical growth or maturity has been achieved. Type B constraints have much more subtle effects on the stages of development and can be imposed on the agent either by external factors, such as levels of stimulation of the environment and interaction with carers, or through non-physical internal mechanisms such as motivations. We investigated these constraints in several experiments using our biologically-based model of eye-head gaze control (Law et al. 2013), in which there is a close relationship between learning eye and head movement behaviours.

In our first experiment we modelled Type A constraints preventing head learning, analogous to a lack of sufficient muscle tone in the neck, and released these at set timed intervals (Law et al., 2013). In a second experiment we compared the effects on learning when a Type A constraint release was triggered by a performance threshold, with the inherent Type B constraint caused by the dependency between eye and head systems (Shaw et al., under review). In our third experiment we investigate the effect of environmental scaffolding as a type B constraint (Shaw et al., 2013).

In Law et al. (2013) we showed that Type A constraints increased the speed of learning in our gaze system. Without any constraints, learning of neck control was slowed by lack of information in the eye system, and eye learning was slowed due to the prolonged motion of the neck. Over the fixed period of the experiment, it was found that the greatest amount of learning in both systems occurred when the constraint was released mid-experiment. The results suggest that the optimal time to release a constraint to maximise learning is dependent on the interaction of the co-dependent learning rates of the systems involved.

In Shaw et al. (under review) we examine further the emergence of the Type B constraint caused by the inherent dependency between eye and head systems. We compare performance to that observed using Type A constraints based on performance and time thresholds, and note that while learning is slower under the Type B constraint, for reasons described above, performance was similar after an equal number of learning cycles. However, the distribution of learnt fields is markedly different, with the Type A constraint resulting in a much denser population of fields in a small region, as opposed to the more sparse, but wider coverage caused by the Type B constraint.

Shaw et al. (2013) show that, when learning with a single visible non-moving stimulus, there is a limitation on the area of the eye mapping that can be learnt by eye movement alone. Movement of the head enables further learning of the eye mapping, but at a reduced rate. By scaffolding learning, by providing more visual stimuli, eye learning can develop further without head involvement (Figure 5.4.1). This demonstrates the environment acting as a Type B constraint.

IM

We have argued that apparently goal-free, motor-based activity is a fundamental process for truly intelligent agents, whether human or robotic, and that this intrinsic activity is the source of behaviours from early motor babbling to play in older children (Lee, 2011). These intrinsic activities were derived from our past experience in building models for developmental robotics which lead us to intrinsic activity as an indispensable component of our algorithms. We developed algorithms for motor babbling and play as intrinsically generated actions, driven by simple novelty, and have shown them to be sufficient for learning complex behaviours.

PSchema

We have developed a framework for embodied agents called PSchema (Sheldon et al., 2011, Sheldon, 2012). Specifically, the framework is designed for Piagetian schema learning and enables the use of symbolic schemas in a robotic environment. We also introduced a generalisation mechanism to further increase the capabilities of schema techniques. We have shown how schemas can be used for long term memory and as a generator for play-like behaviour in robotics. Schemas link pre-conditions, actions, and post-conditions to represent actions and their effects. For instance, a pre-condition could be '*an object is visible at location A*', which links to a known action '*move arm to joint positions x, y*', and finally links to 2 post conditions '*finger in location A*' and '*finger is touching an object*'. Through simple novelty-directed exploration, a robot can learn to generalise schemas with similar conditions, and learn exceptions to these generalisations. Schemas can be chained together to create plans of actions to reach goals, and have been shown to develop new, and surprising, action combinations through novelty-directed learning.

Mapping

Many structures throughout the mammalian nervous system effectively gather their inputs from areas rather than specific points. In doing so, the information can be transmitted or dealt with using far fewer connections than would be needed for the individual points. These areas, knows as fields, usually express smooth curved boundaries that loosely resemble circles or ellipses. Examples can be found in ganglion cells in the retina and in the skin through to deeper cortical areas. Assembled together, these fields form maps of the input space. We have developed a high level mapping technique inspired by these structures and have investigated them as an effective analogy of topological maps in the brain. We model a field as a point in an input space with a radius and have demonstrated this as a content-free biologically-based substrate for sensorimotor learning.

By linking fields that fire at the same time but in separate maps, we can learn to transform a stimulation from one map to another. This allows us to perform tasks such as reaching to areas in the gaze space. Various computational mechanisms can be used to learn connectivity between maps, and we have demonstrated Hebbian learning as a particularly fast and effective method. A key issue relates to the structure and overlap of fields on these maps. The curved nature of fields mean that to fully cover the input space they must overlap and we have explored the properties of these overlaps in detail (Earland et al., Submitted). Through computational analysis, we have shown that optimal field placement and size for robotic applications is similar to that identified in neural structures.

We successfully incorporated weighted vector averaging techniques to combine the mappings of visual, motor configuration and trajectory spaces. These results were transferred from simulation to the robot and combined with compensatory mappings learned on the robot to resolve variations in the two systems. Our work on field structures (Earland et al., 2013) demonstrates the effectiveness, robustness and efficiency of a vector averaging approach to sensorimotor mapping when combined with biologically inspired activation functions. The work also measured the optimisation capabilities of regularly structured mappings over unstructured methods.

Over the course of the project we have developed our framework for representing and navigating a range of sensorimotor spaces including eye, head, torso, arm, and schema memory, and have been extended from simple 2D maps to vector-based representations in multiple dimensions.

Collaboration Efforts

During the project we have collaborated with our IM-CLeVeR partners in a range of activities. In early collaboration efforts we worked to integrate systems from UU and USFD to explore directed search of sensorimotor space.

Later collaboration came with CNR, ISTC and LOCEN on the board experiment. The specific goal was to create a model of learning that will cause a robot to progress through a sequence of stages of development, in a similar way to that of an infant. In (Lee et al. 2012) and (Law et al., Submitted), we describe how we have combined the various systems developed during this project to model stages in infant development, beginning at learning to saccade and ending with open-ended play behaviour, which incorporates torso movements and simple reaching. A further collaboration came with FIAS integrating a visual attention module to identify and learn about objects within the environment.

AU were involved in the ROSSI project on the emergence of communications in robots through sensorimotor and social interaction. As part of this project, AU undertook research into affordances and robot hand grasping. We anticipated transferring this work across to the iCub in year 3 to enable it to better interact with its environment and, therefore, widen the scope for sensorimotor interactions. AU were also involved in the Reverb project on reverse engineering the vertebrate brain. In this work, AU had been collaborating on attention mechanisms and saccade generation. Coordination between the works in these projects and IM-CLeVeR has resulted in a join paper (Huelse et al. 2010).

Throughout much of the project we used the MoBeE collision avoidance software from IDSIA to help ensure the safety of our robot. This software detects impending collisions and then takes over control of the robot, reversing the actions to get the robot back to a safe state. We worked with them further to help us modify the software so that we could get live information about the positions of the robots parts. This was useful for developing sensory motor maps whilst in simulation as the information from the simulator was too inaccurate to transfer to the real robot.

Selected Bibliography

Law, J., Lee, M., Tomasetti, A., Huelse, M., The Infant Development Timeline and its Application to Robot Shaping, Adaptive Behaviour, 2011.

Lee, M., Intrinsic activity: from motor babbling to play, Proceedings of the First Joint International Conference on Development and Learning and Epigenetic Robotics, 1-6, 2011, DOI 10.1109/DEVLRN.2011.6037375

Sheldon, M., Intrinsically motivated developmental learning of communication in robotic agents. PhD thesis, Department of Computer Science, Aberystwyth University, UK., 2012.

Sheldon, M,, Lee, M., PSchema: A developmental schema learning framework for embodied agents, IEEE, Development and Learning (ICDL), IEEE International Conference, 1--7, 2011.

Lee, M., Law, J., Shaw, P., Sheldon, M., An infant inspired model of reaching for a humanoid robot. Proceedings of the IEEE International Joint Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob 2012), San Diego, CA, Nov 2012.

Earland, K., Shaw, P., Law, J., Lee, M. (in Preparation). Overlapping structures in sensorimotor mappings. PLoS ONE.

Law, J., Shaw, P., Lee, M., A biologically constrained architecture for developmental learning of eye–head gaze control on a humanoid robot. Autonomous Robots, 2013. DOI 10.1007/s10514-013-9335-2.

Shaw, P., Law, J., Lee, M., An evaluation of environmental constraints for biologically constrained development of gaze control on an iCub robot. PALADYN Journal of Behavioral Robotics, 3(3) , pp 147-155, 2013,

Huelse, M., McBride, S., Law, J., Lee, M., Integration of active vision and reaching from a developmental robotics perspective, IEEE Transactions on Autonomous and Mental Development, 2(4), 355-367, 2010.

Law, J., Shaw, P., Lee, M., Sheldon, M. (Submitted). From Saccades to Play: A model of Coordinated Reaching through Simulated Development on a Humanoid Robot. Transactions on Autonomous Mental Development.

IDSIA-SUPSI

Our contribution to robotics and machine learning is wide-spread within the IM-CLeVeR project. The different targets of the individual work packages require specific strategies and learning concepts. The Clever-K functional architecture (Pape et al., 2010) is the basis for integrating and demonstrating the individual results as modules in work package 7.

We developed the modular behavioral framework MoBeE (Frank et al., 2012) for the Clever-K architecture. MoBeE is not exclusive applicable for the IM-CLeVeR project; it is designed for any kind of humanoid or other complex robot. However, it depends on the YARP middleware (Metta et al., 2006), which is the software platform for the iCub humanoid robot (Metta et al., 2008). MoBeE contains essential components to apply learning tasks on a real physical robot: It contains a model of the robot for fast forward kinematic simulation and self collision detection and a world model which is used for collision detection and as an abstract representation of the environment. MoBeE is implemented as a filter within the YARP framework. This allows regular YARP modules to communicate with the controlled robot through a transparent safeguard.
We developed a vision module which automatically detects interesting location in the visual field, focussing on objects and trains automatically robust representations of the detected objects (Leitner et al., 2012). Interesting locations of the scene are selected by saliency maps. The objects are detected with a feature detector on the stereo image and the robust tracking of the objects in based on cartesian genetic programming. The detected objects can be joined to the object data base and added to the MoBeE world model.
We developed a robot motion control module which learns task-relevant roadmaps based on natural evolution strategies (Stollenga et al., 2013). A task is a short-term manipulation action like ‘reaching’, ‘pushing sideways’, ‘pushing forward’, and os on. The tasks are defined as constrains with simple mathematical equations in the euclidean work space for specific control points. A control point can be an end-effector of another part of the robot, e.g. the elbow. The learning algorithm finds solutions in the 41 dimensional join space which are homogeneously distributed in the 6 dimensional euclidean work space.
We developed a curiosity-driven autonomous system for learning perceptual invariances and subsequent skills, called Curious Dr. MISFA (Luciw et. al, 2013; Kompella et. al, 2012b) that learns from high-dimensional raw image data, generated from the eyes of an exploring iCub robot. Curious Dr. MISFA enables the iCub to continually learn skills – toppling an object leads to grasping the object, which finally leads to pick & place — starting with no knowledge of its environment, except for a compressed joint-space representation, previously learned by natural evolution (paper in preparation). Through CD-MISFA, the robot explores with the goal to acquire perceptual invariances from Slow Feature Analysis (Wiskott and Sejnowski, 2002), incrementally from the video data. To this end, we use our incremental version of Slow Feature Analysis (IncSFA; Kompella et. al, 2012a) and the version incorporating autoencoders for advanced non-linear processing (AutoIncSFA; Kompella et. al, 2011).
We have extended the efficient coding hypothesis for learning sensory representations to active perception. To this end we have combined sparse coding approaches with a form intrinsically motivated reinforcement learning that favors movements of the sense organs that aid in encoding the sensory data more efficiently. We have demonstrated on the iCub robot how this leads to self-calibrating systems for binocular vision and vergence control (Zhao et al. 2012, Lonini et al., accepted) as well as motion perception and tracking behavior.
We have developed a curiosity-driven vision system that learns to represent and recognize objects in its environment. It utilizes an attention mechanism that drives the system to look to those locations in the environment where it estimates the highest learning progress. The system has been demonstrated on the iCub robot (Chandrashekhariah et al., 2013).
We also focused on the development of novelty detection methods based on biological learning and habituation, developing a new approach for segmentation of objects using live streams from cameras, using a 3D approach that is capable of detecting features and recognizing objects and a new novelty detector learner based on the biological non-associative learning form of habituation, validating the models experimentally with physical robots. We extended core novelty detection methods such that we demonstrated a robot system capable of continuously and autonomously exploring, learning and identifying novel objects within its perceptual and search space using visual processing and habituation. We developed a novel learning architecture using an expandable bag-of-words for effective cumulative learning of visual perceptions.

In the latter years of the project we explored robot interactions with objects and action learning. Action learning methods were further developed based on the previously developed expandable bag-of-word methods incorporating biologically inspired novelty detection for effective exploration and continuous learning. A fuzzy neural network was used to learn and optimise basic affordances through interactions with objects.
Methods were developed to automate composition and parameterization of skills, and learning and adaptation of skills based on a novel evolutionary algorithm. Preliminary work on integration of tactile and visual

modalities to provide additional information related to weight, compressibility, surface texture, temperature and complicated contours in the exterior shape of an object was performed.

In the final year of the project we focused on skill learning, a progression from static to dynamic novelty detection and in particular the incorporation and extension of a neuroscience based model of intrinsic motivation into a practical robotic environment. We implemented two extensions of the Baldassarre et al's intrinsic motivation model (2012) on a PR2 robot; the first extension to the model involved incorporating a probabilistic biased selection approach (PBS) based on former acquired knowledge. The second extension involved predictive learning over time. For this work we devised an experimental setup for action learning where the robot interacted with balls on a table, with holes as targets, and a limited selection of dummy actions. Vision modules enabled the robot to locate an object and use intrinsic motivations to learn to focus and track it. The PR2 robot was able to learn representations for various objects based on intrinsic motivations. The new integrated approach showed consistent improved behavior and clear benefits.

Selected Bibliography

Baldassarre, G., Mannella, F., Fiore, V. G., Redgrave, P., Gurney, K., & Mirolli, M. (2012). Intrinsically motivated action-outcome learning and goal-based action recall: a system-level bio-constrained computational model. *Neural Networks, 41*, 168-187.

Chandrashekhariah, P., Spina, G., Triesch, J. (2013). Let it learn: A Curious Vision System for Autonomous Object Learning. *Int. Conf. on Computer Vision Theory and Applications (VISAPP)*, 2013.

Frank, M., Leitner, J., Stollenga, M., Kaufmann, G., Harding, S., Förster, A., Schmidhuber, J. (2012). The Modular Behavioral Environment for Humanoids & other Robots (MoBeE). *9th International Conference on Informatics in Control, Automation and Robotics (ICINCO)*. Rome, Italy.

Kompella, V. R., Luciw, M., Stollenga, M., Pape, L., Schmidhuber, J. (2012). Autonomous Learning of Abstractions using Curiosity-Driven Modular Incremental Slow Feature Analysis (Curious Dr. MISFA). *IEEE International Conference on Developmental and Learning and Epigenetic Robotics (ICDL-EpiRob)*, San Diego, 2012.

Kompella, V. R., Luciw, M., Schmidhuber, J. (2012). Incremental Slow Feature Analysis: Adaptive Low-Complexity Slow Feature Updating from High-Dimensional Input Streams. *Neural Computation, 24*(11), 2994-3024.

Kompella, V. R., Pape, L., Masci, J., Frank, M., Schmidhuber, J. (2011). AutoIncSFA and Vision-based Developmental Learning for Humanoid Robots. *11th IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, Bled, Slovenia.

Lonini, L., Zhao, Y., Chandrashekhariah, P., Shi, B.E., Triesch, J. (2013). Autonomous Learning of Active Mulit-Scale Binocular Vision. *IEEE Int. Conf. on Development and Learning and Epigenetic Robotics*, accepted.

Luciw*, M., Kompella*, V.R., Kazerounian, S., Schmidhuber, J. (2013). An Intrinsic Value System for Developing Multiple Invariant Representations with Incremental Slowness Learning. *Frontiers in Neurorobotics, 7*(9). *Joint First Authors.

Leitner*, J., Chandrashekhariah*, P., Harding*, S., Frank, M., Spina, G., Förster, A., Triesch, J., Schmidhuber, J. (2012). Autonomous Learning Of Robust Visual Object Detection And Identification On A Humanoid. *IEEE Conference on Development and Learning (ICDL) and Epigenetic Robotics (EpiRob)*. San Diego, USA. November 2012. *Joint First Authors.

Metta, G., Fitzpatrick, P., Natale, L. (2006). YARP: Yet Another Robot Platform. International. *Journal on Advanced Robotics Systems, 3*(1), 43–48.

Metta, G., Sandini, G., Vernon, D., Natale, L., Nori, F. (2008). The iCub humanoid robot: an open platform for research in embodied cognition. *Proceedings of the 8th Workshop on Performance Metrics for Intelligent Systems (PerMIS)*. Gaithersburg, Maryland, USA.

Pape, L., Ring, M., Gomez, F., Frank, M., Förster, A., Migliore, D., Schmidhuber, J. (2010). IM-CLeVeR-K2 functional architecture, *internal report of the EU-funded Integrated Project "IM-CLeVeR* – Intrinsically Motivated Cumulative Learning Versatile Robots".

Stollenga, M., Pape, L., Frank, M., Leitner, J., Förster, A., Schmidhuber, J. (2013). Task-Relevant Roadmaps: A Framework for Humanoid Motion Planning. under review.

Wiskott, L., Sejnowski, L. (2002). Slow feature analysis: Unsupervised learning of invariances. *Neural Computation, 14*(4), 715–770.

Zhao, Y., Rothkopf, C.A., Triesch, J., Shi, B.E. (2012). A Unified Model of the Joint Development of Disparity Selectivity and Vergence Control. *IEEE Int. Conf. on Development and Learning and Epigenetic Robotics*. San Diego, USA. November 2012.